Birth-and-Death Long-Term Evolution Promotes Histone H2B Variant Diversification in the Male Germinal Cell Line

Rodrigo González-Romero,[1] Ciro Rivera-Casas,[1] Juan Ausió,[2] Josefina Méndez,[1] and José M. Eirín-López*[,1]

[1]XENOMAR-CHROMEVOL Group, Departamento de Biología Celular y Molecular, Universidade da Coruña, Coruña, Spain

[2]Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada

***Corresponding author:** E-mail: jeirin@udc.es.

Associate editor: Helen Piontkivska

Research article

Abstract

The rich diversity within each of the five histone families (H1, H2A, H2B, H3, and H4) can hardly be reconciled with the notion of homogenizing evolution. The prevalence of birth-and-death long-term evolution over concerted evolution has already been demonstrated in the linker histone H1 family as well as for the H2A, H3, and H4 core histone families. However, information about histone H2B is lacking. In the present work, we have analyzed the diversity of the members of this histone family across different eukaryotic genomes and have characterized the mechanisms involved in their long-term evolution. Our results reveal that, quite in contrast with other histones, H2B variants are subject to a very rapid process of diversification that primarily affects the male germinal cell lineage and involves their functional specialization probably as a consequence of neofunctionalization and subfunctionalization events after gene duplication. The overall parallelism observed between the molecular phylogenies and the relationships among the electrostatic potentials of the different variants suggests that the latter may have played a major structural selective constraint during H2B evolution. It thus seems that the reorganization of chromatin structure during spermiogenesis might have affected the evolutionary constraints driving histone H2B evolution, leading to an increase in diversity.

Key words: chromatin, H2B variants, selective constraints, spermiogenesis.

Introduction

In contrast to the notion of divergent evolution, most multigene families were thought to be subject to concerted evolution, a process in which a mutation occurring in a repeat spreads all through the gene family members by recurrent unequal crossover or gene conversion (Arnheim 1983). This idea was further reinforced by the general view that a gene family producing a large amount of products needs to maintain a homogeneity among its members (Kedes 1979; Coen et al. 1982; Matsuo and Yamazaki 1989; Thatcher and Gorovsky 1994). During the last four decades, histones have been used (together with ribosomal DNA) to showcase archetypal examples of multigene families subject to concerted evolution. However, based on the high diversity and functional differentiation exhibited by the members of the histone family (Nei and Rooney 2006), the notion of this mechanism representing the major mode of long-term evolution of these proteins has been recently abandoned (Piontkivska et al. 2002; Rooney et al. 2002; Eirín-López et al. 2004a; González-Romero et al. 2008). It has now been clearly demonstrated that the observed variation within the histone family can be better described by a birth-and-death model of evolution that promotes genetic diversity based on recurrent gene duplication events and under a strong purifying selection acting at the protein level (Nei and Hughes 1992), which would eventually lead to the functional differentiation of the new gene copies through a process of neofunctionalization or subfunctionalization (Lynch and Force 2000).

The genetic diversity of the histone family members has critical implications for the function of the nucleosome in different chromatin settings (van Holde 1988) while maintaining the interactions between core histones necessary to build the protein core around which the DNA is wrapped. The latter involves the formation of H2A–H2B and H3–H4 dimers through different protein–protein interactions including those of electrostatic nature. When looking at the diversity within these four core histone families, it seems that although one of each interacting partners is allowed to have a higher extent of variation (H2A and H3), the other maintains a conserved structure (H2B and H4). This is probably important in order to preserve a functional quaternary structure of the nucleosome core particle, able to efficiently bind and package the DNA, as well as to mediate different dynamics process in chromatin metabolism (Ausió 2006).

In this way, the H2B family stands out among histones because of the low extent of diversification of its members (in comparison with members of the H1, H2A, and H3 families) that lack specialized replication-independent variants in the somatic cell lineage, as well as because of the presence of few variants exclusively restricted to the male germinal cell lineage. In humans, two testis-specific variants have been described so far, including a testis-specific H2B (TH2B, also referred to as hTSH2B; Zalensky et al. 2002) and the testis-specific H2B member W (H2BFW, also known as H2BFWT; Churikov et al. 2004). TH2B is specific from testis and from sperm cells

and is encoded by a gene located on chromosome 6 encoding a protein that has 85% and 93% similarity to the canonical H2B and to the testis-specific variant (TH2B) from rat (Kim et al. 1987) and mouse (Choi et al. 1996), respectively. This variant decreases the stability of the histone octamer without compromising its ability to form nucleosomes (Li et al. 2005). On the other hand, histone H2BFW represents a highly divergent variant encoded by a single gene in the X chromosome exclusively transcribed in testis (Churikov et al. 2004). It has 45% and 70% sequence similarity to canonical H2B and TH2B, respectively. This H2B variant is functionally related to the reorganization of chromatin in late stages of spermatogenesis and to the formation of the telomere-binding complex in the human sperm (Gineitis et al. 2000).

Other H2B variants with a lower extent of similarity with canonical H2Bs include subH2Bv, a sperm-specific histone identified in the bull *Bos taurus*, which plays a fundamental role in the development of the mammalian sperm head related to acrosome formation (Aul and Oko 2002); gH2B, a divergent H2B protein identified in *Lilium longiflorum* involved in the packaging of chromatin in pollen (Ueda and Tanaka 1995), which, like protamines, is involved in the remodeling of the male sperm chromatin (Ueda et al. 2000); and H2BV, a variant first identified in *Trypanosoma brucei* that specifically dimerizes with H2A.Z. H2BV exhibits approximately 38% similarity to the canonical H2B and has orthologs in other kinetoplastids (Lowell et al. 2005). In addition, two novel H2B variants involved in pericentric heterochromatin reprogramming during mouse spermiogenesis, referred to as H2BL1 and H2BL2, have been recently identified (Govin et al. 2007), showing resemblance to subH2Bv and H2BFW, respectively.

The prevalence of the birth-and-death mechanism over concerted evolution has already been demonstrated in the linker histone H1 family (Eirín-López et al. 2004a) as well as for H2A, H3, and H4 core histone families (Piontkivska et al. 2002; Rooney et al. 2002; González-Romero et al. 2008). However, information about the diversity and the evolution of H2B is lacking. In the present work, we have analyzed the diversity of H2B family members across eukaryotes, characterizing the mechanisms involved in their long-term evolution. Our results reveal the presence of an incipient process of genetic diversification, resulting from a birth-and-death process under strong purifying selection acting at the protein level, which is focused toward the preservation of a biased amino acid composition in H2B proteins, primarily in the male germinal cell lineage. This report completes a series of previous works aimed to characterize the overall molecular mechanisms driving the evolution of linker and core histone evolution and the functional differentiation of their variants. This information has important implications for future studies on the evolution of chromatin dynamics in the light of the "histone code" (Strahl and Allis 2000; Jenuwein and Allis 2001).

Materials and Methods

Genome Data Mining and Molecular Evolutionary and Phylogenetic Analyses of Histone H2B Genes

We have included in the present analyses all the nonredundant nucleotide sequences encoding H2B histones listed in the NHGRI/NCBI Histone Sequence Database (Marino-Ramirez et al. 2006). In addition, we have performed extensive data-mining experiments in the GenBank database in order to complete and actualize the set of H2B sequences on which this work is based (see supplementary table S1, Supplementary Material online). Multiple sequence alignments were conducted on the basis of the translated amino acid sequences and edited for potential errors using the BIOEDIT (Hall 1999) and CLUSTAL W (Thompson et al. 1994) programs, consisting a total of 155 sequences belonging to 77 different species and encompassing 711 nucleotide sites. The corresponding protein alignment consisted of 147 sequences (due to the presence of eight pseudogenes) showing 237 amino acid positions. The GenBank database and complete genome databases were also screened for the presence of H2B pseudogenes using the Blast tool (Altschul et al. 1990), identifying five human H2B pseudogenes (*Homo sapiens* Ψ.1, Ψ.2, Ψ.3, Ψ.4, and Ψ.5), two mouse H2B pseudogenes (*Mus musculus* Ψ.6 and Ψ.7), and one sea urchin H2B pseudogene (*Strongylocentrotus purpuratus* Ψ.8). The presence of truncated or incomplete H2B sequences, indels in conserved regions, as well as the absence or interruption of the major promoter elements were interpreted as pseudogenization features.

Molecular evolutionary analyses were performed using the computer program MEGA version 4 (Tamura et al. 2007). The extent of nucleotide and amino acid sequence divergence was estimated by means of the uncorrected differences (p distances), as this approach is known to give better results specially for distantly related taxa owing to its smaller variance (Nei and Kumar 2000). The numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site were computed using the modified method of Nei–Gojobori (Zhang et al. 1998), providing the transition/transversion ratio (R) for each case. Evolutionary distances were calculated using the pairwise deletion option in all cases with the exception of the protein phylogenetic inference, where the complete deletion option was used. Standard errors of the estimations were calculated using the bootstrap method (1,000 replicates).

The phylogeny of H2B proteins was reconstructed using the maximum-likelihood method. The model of protein evolution that best fits the set of H2B sequences analyzed in the present work was selected by using the ProtTest ver. 2.4 program (Abascal et al. 2005) using the Akaike information criterion, defining the Whelan And Goldman model (Goldman and Whelan 2001) as the most fitted to the data. Maximum-likelihood phylogenetic trees were reconstructed with PhyML (Guindon and Gascuel 2003). To this end, the proportion of invariable sites and the shape of the gamma parameter were estimated. The reliability of the

optimized topology obtained was assessed using the approximate likelihood ratio test (Anisimova and Gascuel 2006) and nonparametric bootstrap analysis (100 replicates). Phylogenetic trees were additionally reconstructed using the neighbor-joining tree-building method (Saitou and Nei 1987). The reliability of the resulting topologies was tested by the bootstrap method (Felsestein 1985) and by the interior-branch test (Sitnikova 1996), producing the bootstrap probability and confidence probability values for each interior branch in the tree, respectively. An estimate of the rates of evolution for H2B protein variants was determined using divergence data from the analyses in the present work. Divergence times between taxa were defined according to Hedges and Kumar (2009) as detailed in supplementary table S2, Supplementary Material online.

Gauge of Selection and Nature of the Selective Constraints Acting on H2B Histones

The presence and nature of selection on H2B histones was studied using four different complementary approaches: First, the codon usage bias in H2B histones was referred to as the effective number of codons (Wright 1990) and was estimated using the program DnaSP version 5 (Librado and Rozas 2009). Second, the codon-based *Z* test for selection was used to compare the numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site, defining 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). The *Z* statistic and the probability of rejecting the null hypothesis were obtained. Third, the influence of selection on certain overrepresented amino acids was analyzed by determining the correlation between the genomic content of Guanine and Cytosine nucleotides (GC content, estimated at 4-fold degenerate sites; Li 1997) and the proportion of GC-rich (glycine, alanine, proline, and trytophan [GAPW]) and GC-poor (phenylalanine, tyrosine, methionine, isoleucine, and lysine [FYMINK]) amino acids. Under the neutral model, GC-rich and GC-poor amino acids will be positively and negatively correlated with genomic GC content, respectively (Kimura 1983; Jukes and Bhushan 1986). Correlations were computed using the Spearman rank correlation coefficient, and statistical significance was assessed using standard regression analyses. Fourth, the effect of mutation and selection bias at the nucleotide level was studied by comparing nucleotide frequencies at first codon positions (always nonsynonymous in the case of the residues studied here) and at 4-fold codon positions (always synonymous). Nucleotide frequencies should not be significantly between both positions under the neutral model (Kimura 1983).

Tertiary Structure Modeling and Reconstruction of Electrostatic Potentials of H2B Molecules

The tertiary structures of the H2B proteins were modeled using the coordinates determined for the crystal structure of the H2B histone from *Xenopus laevis* (Protein Data Bank accession code 1AOI) as a reference (Luger et al. 1997) in the context of the SWISS-MODEL workspace (Arnold et al. 2006). The obtained structures were rendered using the MacPyMOL program (DeLano 2007). Comparisons between the electrostatic properties of H2B 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

Phylogenetic and Molecular Evolutionary Study of the Long-Term Evolution of the Histone H2B Family

The first approach to characterize the long-term evolution of the members of the H2B family involved the reconstruction of the phylogenetic relationships using the protein and gene sequences available for the members of this histone family. The phylogenetic tree based on protein sequences exhibits a topology in which the different taxonomic groups are well defined (fig. 1 and supplementary fig. S1, Supplementary Material online), and where it is possible to discriminate a clustering pattern of the different types (irrespectively of the species to which they belong). Under a concerted evolution model, the homogenization of DNA sequences would also result in a certain balance in the nature of the nucleotide substitutions observed among the family members, with a lack of significant differences between synonymous and nonsynonymous substitutions. Furthermore, the divergence of inactive H2B pseudogenes respective to H2B functional genes will be prevented, given that unequal crossing over or gene conversion would homogenize DNA segments as blocks regardless of the functionality of the genes included. By contrast, birth-and-death would generate new gene copies ("birth" process) subject to a subsequent process of strong purifying selection acting at the protein level. If one of the copies becomes inactive (pseudogenized, "death" process), it could remain in the genome for long periods of time but excluded from the action of selection at the protein level. As a result, pseudogenes would evolve following random genetic drift becoming highly divergent copies respect to functional genes (Eirín-López et al. 2004a; Nei and Rooney 2006). With this in mind, an analysis of the nucleotide variation among H2B genes was carried out. The results revealed a high extent of synonymous divergence (leading to the saturation level in most comparisons as depicted by supplementary fig. S2, Supplementary Material online) that is significantly greater than the nonsynonymous variation in all comparisons (table 1). Indeed, p_S acquires closely related values within and between species comparisons, suggesting that intraspecific homogenization does not play a major role in the long-term evolution of H2B (otherwise, the interspecific divergence would be significantly higher than the intraspecific polymorphism).

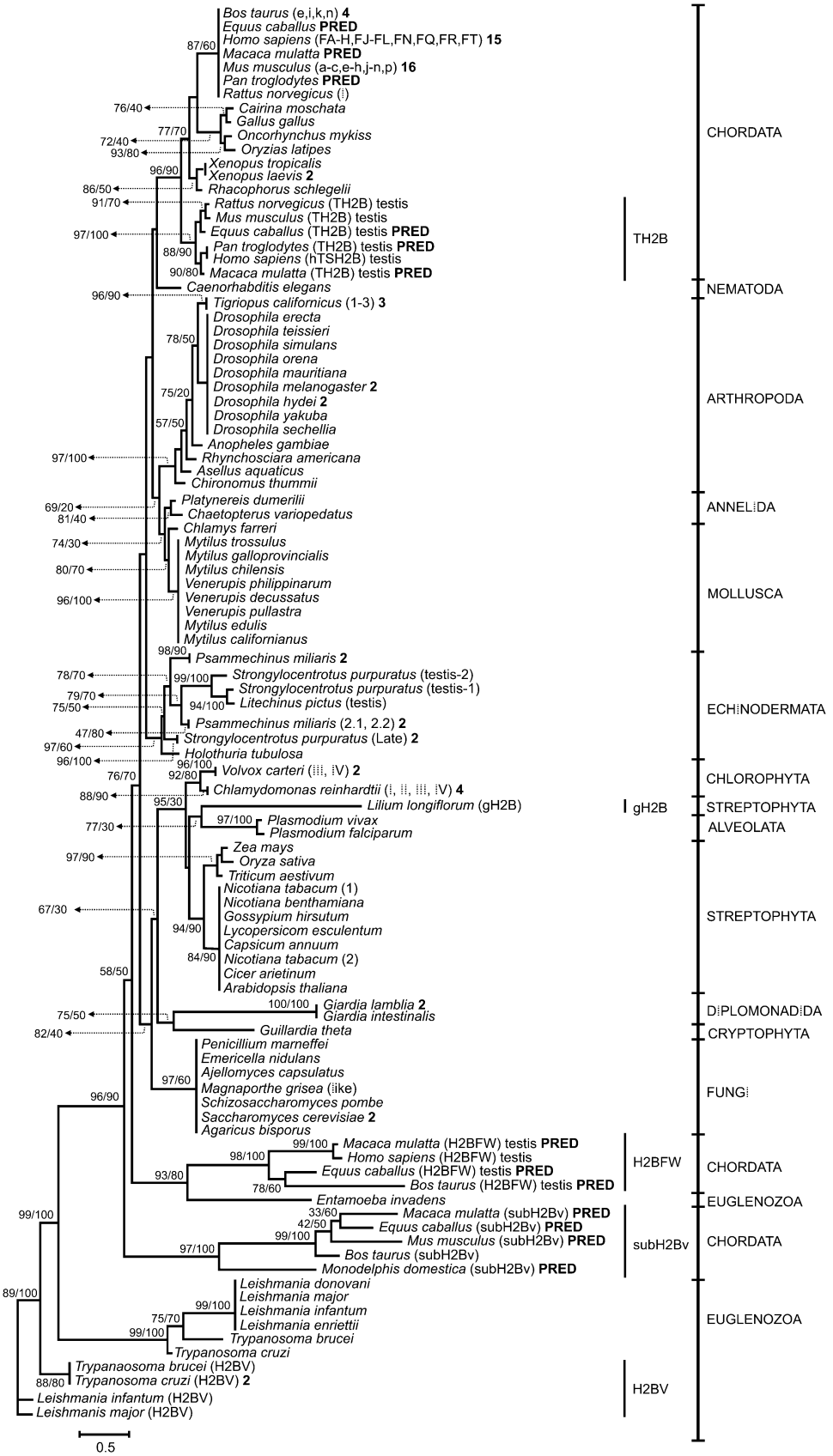

Fig. 1. Phylogeny of H2B proteins from different eukaryotic groups (see supplementary table 1, Supplementary Material online) reconstructed using maximum likelihood. The numbers for interior branches represent approximate likelihood ratio test values followed by nonparametric bootstrap probabilities based on 100 replications and are only shown when at least one of the values is ≥50%. Numbers in parentheses near species indicate the H2B subtype and numbers in boldface represent the number of sequences analyzed for each species. Taxonomic groups are indicated in the right margin of the tree. Sequences predicted from databases and draft genomes data are indicated by predicted near the species name.

Table 1. Average Numbers of Synonymous (p_S) and Nonsynonymous (p_N) Nucleotide Differences per Site and Average Transition/Transversion Ratio (R) in H2B Genes from Representative Protostomes, Deuterostomes, Plants, Protists, and Fungi.

	p_S (SE)	p_N (SE)	R
Deuterostomes			
Human	0.314 (0.020)	0.010 (0.003)	2.0*
Bos taurus	0.302 (0.029)	0.013 (0.005)	1.8*
Mouse	0.201 (0.017)	0.015 (0.004)	1.3*
Xenopus	0.270 (0.030)	0.009 (0.005)	1.6*
Human/*Bos taurus*	0.339 (0.024)	0.010 (0.004)	1.7*
Mouse/rat	0.192 (0.025)	0.029 (0.011)	1.2*
Mammals	0.309 (0.020)	0.014 (0.004)	1.3*
Xenopus/Rhacophorus	0.357 (0.038)	0.023 (0.010)	1.6*
Fish	0.322 (0.048)	0.029 (0.013)	0.7*
Strongylocentrotus purpuratus	0.464 (0.032)	0.168 (0.022)	0.9*
Sea urchins	0.502 (0.024)	0.163 (0.019)	0.9*
Protostomes			
Drosophila	0.227 (0.019)	0.004 (0.002)	2.1*
Drosophila/Rhynchosciara	0.622 (0.043)	0.037 (0.015)	1.8*
Annelids	0.460 (0.056)	0.133 (0.026)	0.7*
Moluscs	0.109 (0.012)	0.021 (0.005)	1.0*
Plants			
Nicotiana tabacum	0.529 (0.047)	0.084 (0.017)	1.7*
Plants	0.617 (0.018)	0.115 (0.014)	1.2*
Protists			
Volvox carteri	0.375 (0.043)	0.038 (0.012)	1.0*
Chlamydomonas reinhardtii	0.101 (0.019)	0.019 (0.007)	1.5*
Volvox/Chlamydomonas	0.402 (0.034)	0.116 (0.022)	0.6*
Fungi			
Saccharomyces	0.333 (0.043)	0.031 (0.014)	2.7*
Fungi	0.594 (0.018)	0.134 (0.016)	1.3*
TH2B	0.482 (0.025)	0.029 (0.008)	2.2*
H2BFWT	0.510 (0.048)	0.373 (0.028)	0.6*
H2BV	0.504 (0.033)	0.124 (0.019)	1.0*
SubH2Bv	0.448 (0.027)	0.297 (0.020)	1.0*

NOTE.—$p_S > p_N$ in Z-test comparisons. SE indicates standard errors calculated by the bootstrap method with 1,000 replicates.
*$P < 0.001$.

In addition to all the above, our analyses identified the presence of up to eight H2B pseudogenes (Ψ) in the genomes from different deuterostomes (see supplementary table S1, Supplementary Material online). Comparisons between the amount of nucleotide variation detected within pseudogenes and that observed between pseudogenes and functional H2B genes revealed the presence of significant differences in five of eight pseudogenes analyzed (three from human, one from mouse, and one from sea urchin), suggesting the absence of a significant effect of gene conversion or interlocus recombination in these instances (table 2). Furthermore, the presence of stretches of shared mutations unambiguously present in repeat variants at exactly the same positions has been proposed to be a consequence of homogeneization driven by gene conversion mechanisms. Therefore, the significance of these observations was analyzed by using the GENECONV program (Sawyer 1999), which revealed a total absence of stretches of homogenized sequences resulting from gene conversion events among canonical and variant H2B representatives.

Given that concerted evolution does not allow the functional differentiation of genes (because all members genes are supposed to evolve as a unit), our results point toward the absence of any significant homogenization process in the testis-specific TH2B variants, the subH2Bv variants from mammals, as well as for the H2BV isoforms from protists. The protein phylogeny suggests that the novel H2BL1 and H2BL2 variants that are involved in pericentric heterochromatin reprogramming during mouse spermiogenesis (Govin et al. 2005) are indeed mouse subH2Bv and H2BFW variants, as indicated by the clustering pattern in the tree. Overall, our results agree with the model of birth-and-death evolution based on the generation of genetic diversity through recurrent gene duplications. According to this model, new H2B copies would be generated and subject to strong purifying selection acting at the protein level. The effect of selection over the gene diversity generated would favor the functional differentiation of H2B subtypes, probably as a consequence of neofunctionalization or subfunctionalization events.

Evolutionary Constraints Behind Histone H2B Evolution

If concerted evolution does not play a major role in the long-term evolution of H2B, the question arises regarding the major constraints involved in H2B diversification and differentiation during evolution. In this regard, the nature of the selective process operating on H2B genes was initially analyzed by comparing the amount of synonymous and nonsynonymous variation among different family members. Our results showed that, in all instances, the synonymous variation is significantly greater than the nonsynonymous variation (table 1). This was also true in comparisons performed both within and between species ($P < 0.001$, Z test). These results are indicative of the presence of strong purifying selection acting at the protein level and aimed at maintaining the relevant structural organization of H2B histones that is required for their function in the nucleosome. The observations from the synonymous/nonsynonymous comparison are nicely complemented by the study of the amount of codon bias exhibited by H2B genes (table 3), which shows that canonical H2B sequences from protostomes are significantly less biased (46.914 ± 7.617) when compared with canonical deuterostome H2B sequences (34.507 ± 5.848; $P < 0.001$, t test).

Given the basic nature of histones and the role of their basic residues in protein–DNA and protein–protein interactions, the selection for certain biased amino acids in H2B variants was investigated in order to ascertain the potential selective constraints leading to their differentiation. The most abundant residues in H2B variant proteins (TH2B, subH2Bv, H2BFW, and H2BV) are represented by alanine (GC rich) and lysine (GC poor). The presence of selection for such amino acids was subsequently analyzed by studying the correlation between the frequency of GC-rich and

Table 2. Pseudogene and Functional H2B Nucleotide Divergences Using *p* Distances.

		Divergence *P* distance (SE)	
Pseudogene	**Evidence[a]**	**Pseudogene versus functional**	**Average functional genes**
Human H2B (Ψ.1)	**FSM255**	**0.344 (0.021)**	**0.099 (0.010)***
Human H2B (Ψ.2)	**STP129**	**0.113 (0.013)**	**0.099 (0.010)**
Human H2B (Ψ.3)	**STP129**	**0.101 (0.010)**	**0.099 (0.010)**
Human H2B (Ψ.4)	**FSM280**	**0.289 (0.020)**	**0.099 (0.010)***
Human H2B (Ψ.5)	**FSM280**	**0.292 (0.020)**	**0.099 (0.010)***
Mouse H2B (Ψ.6)	**FSM380**	**0.045 (0.006)**	**0.064 (0.007)**
Mouse H2B (Ψ.7)	**FSM195**	**0.395 (0.022)**	**0.064 (0.007)***
Sea urchin H2B (Ψ.8)	**DEL200**	**0.185 (0.030)**	**0.062 (0.013)***

NOTE.—Standard errors were computed by the bootstrap method (1,000 replicates) and are indicated in parentheses.

[a] Evidences for pseudogenization are indicated as: FSM, frameshift mutation; DEL, deletion of gene segment; STP, stop codon. Numbers indicate the position at which mutations occur.

*$P < 0.001$ in codon-based *Z*-test comparisons between pseudogene and functional genes.

GC-poor amino acids with the genomic GC content (see supplementary table S3, Supplementary Material online). We did not find a significant correlation between the genomic GC content and the frequency of GC-rich and GC-poor amino acids across H2B variants (fig. 2). Similarly, a significant correlation between genomic GC content and the most represented amino acids in each class (alanine and lysine) was also absent as indicated in figure 2 and table 4 ($P > 0.05$ in all Spearman rank correlations), suggesting a departure from what would be expected from a neutral evolution model.

The implicit presence of selection behind these observations was further assessed by an additional method to gauge the significance of mutation bias and selection at the nucleotide level in the evolution of H2B variants. This involves the comparison of changes at first codon positions (nonsynonymous) with those at third codon positions (synonymous) in the most frequent residues in H2B protein variants. Under the neutral model, the nucleotide frequencies at both positions should not be significantly different. Codons for alanine (GC rich) contain G at first codon positions, whereas codons for lysine (GC poor) have A at first codon positions. Analysis of the mean G + A content at first codon positions in TH2B (66.033 ± 0.403), subH2Bv (56.260 ± 8.307), H2BFW (57.400 ± 2.515), and H2BV (61.200 ± 1.805) showed that their values were significantly larger than the mean G + A content at 4-fold degenerate positions in TH2B (37.917 ± 5.110), subH2Bv (40.880 ± 4.485), H2BFW (33.400 ± 3.653), and H2BV (44.600 ± 9.110) (TH2B, t test = 13.435, $P = 0.000$; subH2Bv, t test = 3.619, $P = 0.006$; H2BFW, t test = 10.822, $P = 0.000$; and H2BV, t test = 13.435, $P = 0.011$). Our results strongly suggest that selection has acted to maintain high levels of alanine and lysine in the different H2B variants in striking contrast to the predictions in a neutral scenario in which amino acid and nucleotide compositions would be driven by the underlying GC content as a result of mutation bias.

Table 3. Codon Usage Bias (ENC, effective number of codons) in H2B Histone Genes and H2B Gene Variants.

Taxonomic group/H2B variant	**ENC**
Deuterostomes	**34.507 ± 5.848**
Protostomes	**46.914 ± 7.617**
TH2B	**43.648 ± 6.497**
H2BFW	**45.101 ± 6.147**
SubH2Bv	**52.829 ± 3.932**
H2BV	**46.987 ± 6.163**
Others	**35.167 ± 8.122**

Electrostatic Potentials and Histone H2B Diversity

Ionic interactions play a fundamental role in the way histone H2B interacts with other histones and with DNA that modulate chromatin dynamics. Therefore, we decided to analyze the potential selection role of the electrostatic interaction properties of the different members of the H2B family. The electrostatic potentials and the corresponding similarity indices were calculated for all H2B histones listed in supplementary table S1, Supplementary Material online. This allowed us to calculate the electrostatic distances between proteins. The epogram shown in figure 3 depicts the electrostatic distances between H2B members in a topology that is in agreement with the protein phylogenies shown in figure 1 and supplementary figure S1, Supplementary Material online. Although canonical H2B proteins from protostomes, deuterostomes, as well as TH2B variants are located within the same group at the upper part of the epogram, H2B histones from plants and protists cluster with the rest of the H2B variants but in clearly different subgroups at the lower side of the epogram. An exception to this topological coincidence is represented by TH2B from *Rattus norvegicus*, which is located in the lower part of the epogram, more closely related to H2BV variants from protists than to its mammalian counterparts.

Discussion

Homogenizing versus Birth-and-Death Evolution in the Histone H2B Family

In the present work, we have focused our attention in the histone H2B family because of the low level of diversification displayed by its members and also because of the complete lack of information on its long-term evolution. The reconstructed topologies showed that different H2B variants are clustered by type and not by species (fig. 1 and supplementary fig. S1, Supplementary Material

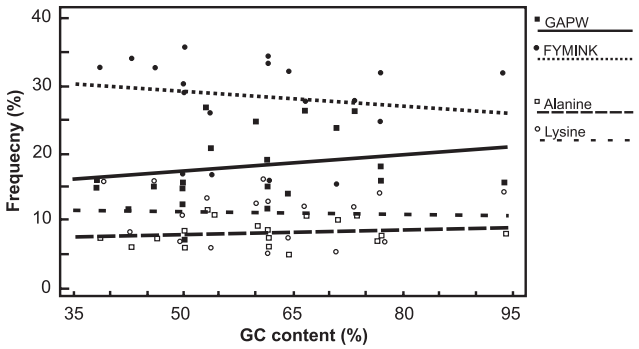

Fig. 2. Relationship between the GC content and the frequencies of GC-rich (GAPW) and GC-poor (FYMINK) amino acid classes and between the GC content and the frequencies of alaline and lysine residues in H2B variants.

online). The extent of the synonymous divergence is always significantly greater than that of the nonsynonymous divergence both within and between species (table 1), suggesting the presence of an extensive silent divergence among H2B genes (supplementary fig. S2, Supplementary Material online). Also, most of the estimated intraspecific p_S values were as high as the interspecific p_S values (table 1). Histone H2B pseudogenes showing significant differences with respect to functional H2B genes could also be detected in human (Ψ.1, Ψ.4, and Ψ.5), mouse (Ψ.7), and sea urchin (Ψ.8), as shown in supplementary fig. S2, Supplementary Material online, and table 2. All these observations provide evidence for a gradual model of evolution in the H2B family based on birth-and-death under strong purifying selection acting at the protein level. Accordingly, the nucleotide divergence among family members will be primarily synonymous and pairs of genes resulting from a recent duplication will be expected to be closely related or even identical (Nei et al. 2000). As with canonical H2B genes, concerted evolution has not had a major effect on the long-term evolution of H2B variants, which is mainly driven by a birth-and-death process (Nei and Hughes 1992).

On the Mechanisms Underlying the Differentiation of H2B Variants

We decided to further investigate the nature of the long-term evolutionary mechanisms of histone H2B using three different approaches. First, comparisons of synonymous and nonsynonymous nucleotide substitutions among H2B members evidenced the presence of extensive silent divergence (supplementary fig. S2, Supplementary Material online, and table 1), with H2B genes from deuterostomes were significantly more biased than their protostome counterparts (table 3). This is despite the fact that differences in codon bias are common between different histone multigene families more or less independently of the particular organisms studied (Eirín-López et al. 2002; Eirín-López et al. 2004b; González-Romero et al. 2008). Such result can be explained by the higher degree of functional specialization that is observed in deuterostome H2B genes when compared with the apparently less differentiated H2B genes from protostomes, as is also the case of histone H1 (González-Romero et al. 2008). In this regard, the low bias displayed by the variant subH2Bv could be accounted for by its role in acrosome formation, outside the sperm nucleus (Aul and Oko 2002). An alternative hypothesis for the codon bias could be related to the preferential use of certain preferred triplets encoding overrepresented amino acids in H2B proteins, such as it is observed for other highly basic proteins (Eirín-López et al. 2006; González-Romero et al. 2008).

Table 4. Correlations Between Genomic GC Content and the Frequencies of GC-Rich (GAPW) and GC-Poor (FYMINK) Amino Acids in H2B Histone Variants.

Histone H2B variants	r_S	P
Genomic GC versus GAPW (GC rich)	0.349	0.127
Genomic GC versus FYMINK (GC poor)	−0.377	0.099
Genomic GC versus alanine	0.205	0.371
Genomic GC versus lysine	−0.265	0.247

NOTE.—r_S is the Spearman rank correlation coefficient.

Second, the study of the frequencies displayed by different amino acids (GC rich and GC poor) revealed a lack of a significant correlation between the frequencies of any of these two classes and the genomic GC content in H2B variants, contrasting with what would be expected from a neutral evolutionary process (Kimura 1983; Jukes and Bhushan 1986), suggesting the existence of a selection mechanism maintaining a biased amino acid composition (fig. 2 and table 4). Given the relevance of the positive charge contributed by the basic amino acids of histones to their electrostatic interactions, the presence of a selection mechanism favoring biased basic residues was further analyzed. Our results pointed again toward a departure from the predictions from a neutral mechanism of evolution for the frequencies of the most abundant amino acids of the H2B variants (fig. 2 and table 4), including the basic amino acid lysine (GC poor) as well as the nonpolar amino acid alanine (GC rich). Therefore, the selection mechanism appears to maintain a biased amino acid composition in H2B variants that affects these two abundant amino acids.

Third, our results revealed significant differences between the changes occurring at first (nonsynonymous) and third (synonymous) codon positions of the most frequent amino acid residues, with a larger difference observed for the first codon positions. Although the neutral model of molecular evolution predicts that amino acid and nucleotide compositions are driven by the underlying GC content as a result of mutation bias (Kimura 1983), our data strongly suggest that selection has acted to maintain high levels of lysine and alanine, biasing the nucleotide composition of H2B variants. Only few studies in addition to the present work have underscored the higher relevance of natural selection over mutation bias in determining amino acid composition of proteins (Rooney et al. 2000; Akashi and Gojobori 2002; Rooney 2003; Eirín-López et al. 2006; González-Romero et al. 2008).

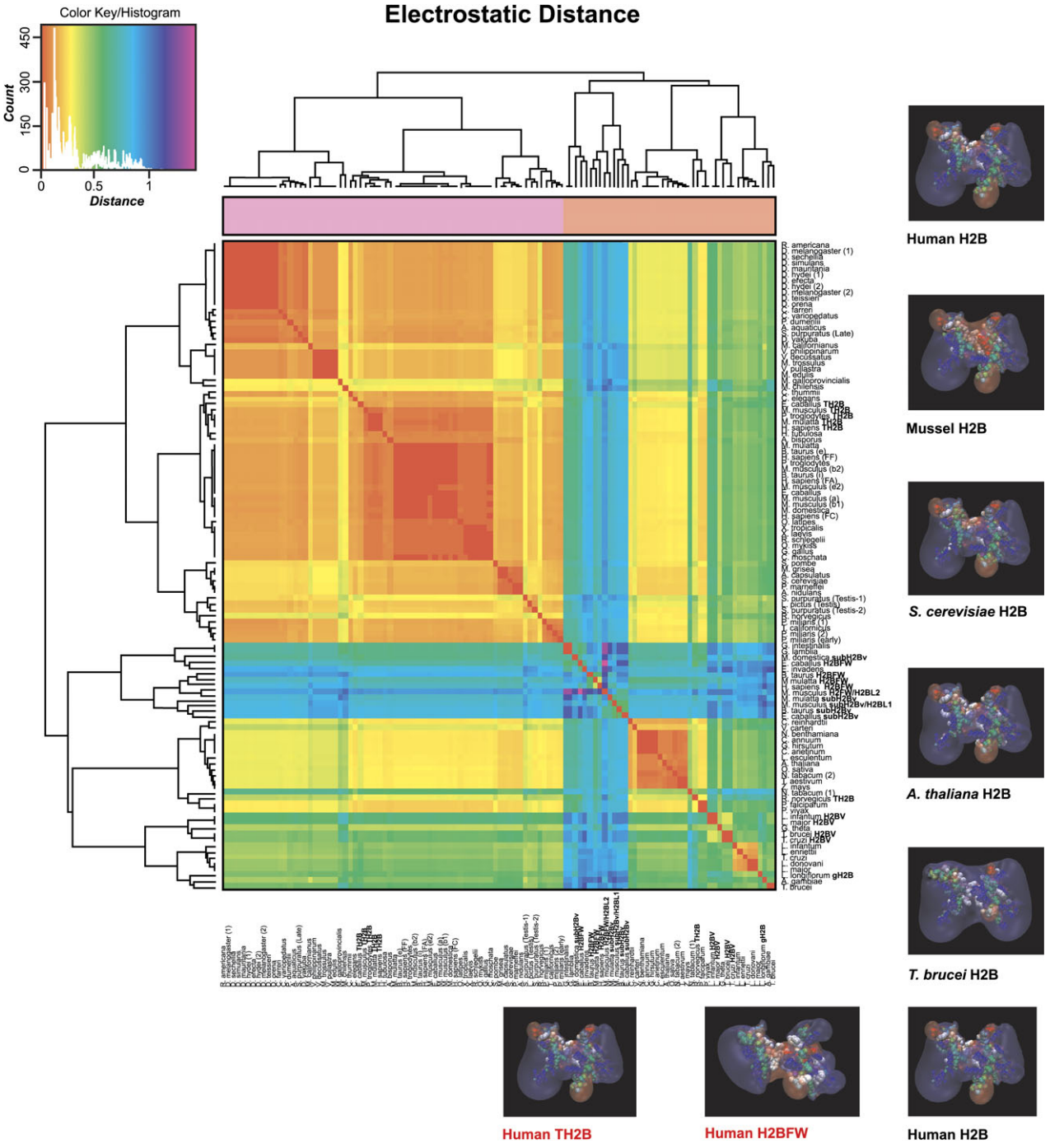

Fig. 3. Electrostatic distances calculated from the similarity indices for the electrostatic potentials of histone H2B proteins represented in a color-coded matrix (heat map). The distance between similarity indices (S) of every pair of molecules (a and b) is defined as $D_{a,b} = \sqrt{2 - 2S_{a,b}}$ (Wade et al. 2001). The color code and the number of comparisons for each distance interval are indicated in the key/histogram. The tree along the side of the image assembles the proteins into groups with similar electrostatic potentials (epogram). The electrostatic potentials of five representative canonical H2B molecules belonging to different taxonomic groups are represented in the right margin of the figure as well as the electrostatic potentials for H2B variants TH2B and H2BFW from human (below the epogram). Negatively and positively charged surfaces are depicted in red and blue, respectively; colors were assigned to amino acids in the reconstruction of the 3D structures according to their physical and chemical structural characteristics (red, acid; blue, basic; green, polar uncharged; and purple, nonpolar hydrophobic).

Electrostatic Interactions Impart a Potential Constraint to Histone H2B Differentiation

Electrostatic interactions play an important role in histone–histone and histone–DNA interactions involved in chromatin dynamics and metabolism. The epogram shown in figure 3 distinctively discriminates between the different groups of H2B variants analyzed, providing a topology that is in agreement with the phylogenetic

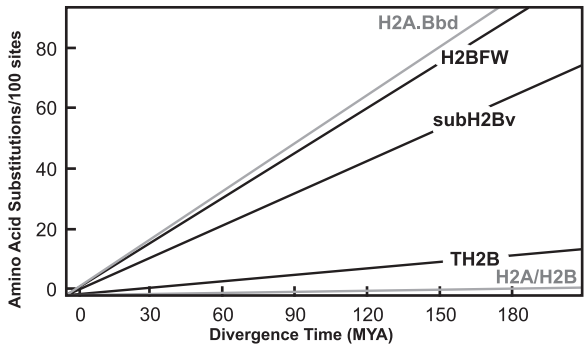

Fig. 4. Evolutionary rates (amino acid change/100 sites) for the H2B variants studied in the present work compared with those from the fast-evolving mammalian histone H2A.Bbd (Eirín-López et al. 2008; Ishibashi et al. 2010) and from the slow-evolving H2A/H2B proteins for comparison.

relationships reconstructed from other sequence information in the present work. This analysis clearly discriminates between the group encompassing the divergent H2BFW and subH2Bv variants and the remainder of the H2B histones from animals (where the TH2B group is included) and plants. Notwithstanding, an exception occurs with TH2B from rat, which is located in the lower part of the epogram, appearing to be more closely related to H2BV variants from protists than to the mammalian counterpart (as it is also observed in the protein phylogeny). Such disagreement is most likely the result of an amino acid change from lysine, a basic residue, to glutamic acid at amino acid position 36 of the protein, emphasizing once more, the relevance of protein variation in the context of the functional differentiation of H2B variants. The overall parallelism observed between the molecular phylogenies and the epogram suggests that electrostatic potentials represent an important structural constraint that is subject to selection during H2B evolution. It thus appears that modifications in key residues affecting the overall electrostatic potential could constitute the mechanistic basis for the functional differentiation of H2B variants as it has been previously reported for histone H2A (Eirín-López et al. 2009).

H2B Diversity and the Germinal Cell Line

The constraints driving the long-term evolution of the H2B family involve the maintenance of high levels of certain biased amino acids (lysine and alanine), which are important for the establishment of the correct interactions involved in the formation of the nucleosome. Quite in contrast with other histones, H2B members are subject to a very rapid process of diversification that is primarily restricted to the male germinal cell lineage and involves the functional specialization of different histone variants probably as a consequence of neofunctionalization and subfunctionalization events after gene duplication. This is specifically evident in the case of H2BFW (see fig. 4) that evolves almost at the same rate as the quickly evolving histone H2A.Bbd that is also involved in mammalian spermiogenesis (Eirín-López et al. 2008; Ishibashi et al. 2010).

The lack of diversity within the H2B and H4 families has been regarded to be the result of their essential role in the maintenance of the fundamental structural H2A–H2B and H3–H4 domains of the nucleosome. By contrast, the variation presented by the H2A and H3 counterparts would be responsible for imparting different functional and structural specificities to these domains (Ausió 2006). Such a hypothesis would be consistent with the increase in H2B diversity observed in the male germinal cell line where a dramatic change in chromatin conformation takes place during spermiogenesis. Two conclusions can be drawn from this: First, H2B variation implicitly suggests the possibility of H4 variation. Indeed, the few H4 variants described to date are mostly circumscribed to testis (Grimes et al. 1987; Wolfe et al. 1989; Wolfe and Grimes 1991). Second, the diversification of H2B and H4 histones would be absent from the female germinal cell line (i.e., in oocytes) due to the prevalence of a nucleosome chromatin organization, which would only be compatible with H1 variants such as H1oo (Tanaka et al. 2001) and H1M/B4 (Cho and Wolffe 1994). It thus seems that the reorganization of chromatin structure during spermiogenesis might have affected the evolutionary constraints driving histone H2B evolution, leading to an increase in diversity. However, with the exception of a few structural studies (Kim et al. 1987; Choi et al. 1996; Zalensky et al. 2002; Li et al. 2005), little is known about the specific role performed by the testis-specific H2B variants. Further studies will be needed in order to clearly decipher the connection between the relaxation of the evolutionary constraints described here and the drastic structural chromatin transitions involved in spermiogenesis.

Supplementary Material

Supplementary figures S1 and S2, tables S1–S3, and other materials are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by a contract within the Isidro Parga Pondal Program (Xunta de Galicia) and by a contract within the Ramon y Cajal Subprogramme from the Spanish Government-MICINN (to J.M.E.-L.); by a grant from the Xunta de Galicia (07MMA013103PR to J.M.); 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. C.R.-C. has been funded by a collaboration research fellowship from the Spanish Government-MEC.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

Akashi H, Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A*. 99:3695–3700.

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

Anisimova M, Gascuel O. 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Syst Biol*. 55:539–552.

Arnheim N. 1983. Concerted evolution of multigene families. In: Nei M, Koehn RK, editors. Evolution of genes and proteins. Sunderland (MA): Sinauer Associates. p. 38–61.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201.

Aul RB, Oko RJ. 2002. The major subacrosomal occupant of bull spermatozoa is a novel histone H2B variant associated with the forming acrosome during spermiogenesis. *Dev Biol*. 242: 376–387.

Ausió J. 2006. Histone variants: the structure behind the function. *Brief Funct Genomic Proteomic*. 5:228–243.

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

Choi YC, Gu W, Hecht NB, Feinberg AP, Chae CB. 1996. Molecular cloning of mouse somatic and testis-specific H2B histone genes containing a methylated CpG island DNA. *Cell Biol*. 15: 495–504.

Churikov D, Siino J, Svetlova M, Zhang K, Gineitis A, Morton Bradbury E, Zalensky AO. 2004. Novel human testis-specific histone H2B encoded by the interrupted gene on the X chromosome. *Genomics* 84:745–756.

Coen E, Strachan T, Dover GA. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species subgroup of *Drosophila*. *J Mol Biol*. 158:17–35.

DeLano WL. 2007. MacPyMOL: a PyMOL-based molecular graphics application for MacOS X. Palo Alto (CA): DeLano Scientific LLC.

Eirín-López JM, Frehlick LJ, Ausio J. 2006. Long-term evolution and functional diversification in the members of the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Genetics* 173:1835–1850.

Eirín-López JM, Gonzalez-Romero R, Dryhurst D, Ishibashi T, Ausio J. 2009. The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues. *BMC Evol Biol*. 9:31.

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, Ishibashi T, Ausió J. 2008. H2A.Bbd: a quickly evolving hypervariable mammalian histone that destabilizes nucleosomes in an acetylation-independent way. *FASEB J*. 22:316–326.

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.

Felsestein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.

Gineitis A, Zalenskaya I, Yau P, Bradbury EM, Zalensky AO. 2000. Human sperm telomere-binding complex involves histone H2B and secures telomere membrane attachment. *J Cell Biol*. 151:1591–1598.

Goldman N, Whelan S. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18:681–699.

González-Romero R, Ausió J, Méndez J, Eirín-López JM. 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*. 66:505–518.

Govin J, Caron C, Rousseaux S, Khochbin S. 2005. Testis-specific histone H3 expression in somatic cells. *Trends Biochem Sci*. 30:357–359.

Govin J, Escoffier E, Rousseaux S, et al. (11 co-authors). 2007. Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol*. 176:283–294.

Grimes S, Weisz-Carrington P, Daum H, Smith J, Green L, Wright K, Stein GS, Stein JL. 1987. A rat histone H4 gene closely associated with the testis-specific H1t gene. *Exp Cell Res*. 173:534–545.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:592–704.

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.

Hedges SB, Kumar S. 2009. The timetree of life. New York: Oxford University Press.

Humphrey W, Dalke A, Schulten K. 1996. VMD—visual molecular dynamics. *J Mol Graph*. 14:33–38.

Ishibashi T, Li A, Eirín-López JM, Zhao M, Missiaen K, Abbott DW, Meistrich ML, Hendzel MJ, Ausió J. 2010. H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis. *Nucleic Acids Res*. 38:1780–1789.

Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* 293:1074–1080.

Jukes TH, Bhushan V. 1986. Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. *J Mol Evol*. 24:39–44.

Kedes L. 1979. Histone genes and histone messengers. *Annu Rev Biochem*. 225:501–510.

Kim YJ, Hwang I, Tres LL, Kierszenbaum AL, Chae CB. 1987. Molecular cloning and differential expression of somatic and testis-specific H2B histone genes during rat spermatogenesis. *Dev Biol*. 124:23–34.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Li WH. 1997. Molecular evolution. Sunderland (MA): Sinauer.

Li A, Maffey AH, Abbott WD, Conde e Silva N, Prunell A, Siino J, Churikov D, Zalensky AO, Ausio J. 2005. Characterization of nucleosomes consisting of the human testis/sperm-specific histone H2B variant (hTSH2B). *Biochemistry* 44:2529–2535.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Lowell JE, Kaiser F, Janzen CJ, Cross GA. 2005. Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in *Trypanosoma brucei*. *J Cell Sci*. 118:5721–5730.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389:251–260.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.

Madura JD, Briggs JM, Wade RC, et al. (11 co-authors). 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. *Comput Phys Commun*. 91:57–95.

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.

Matsuo Y, Yamazaki T. 1989. Nucleotide variation and divergence in the histone multigene family in *Drosophila melanogaster*. *Genetics* 122:87–97.

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, editors. 11th Histocompatibility Workshop and Conference. Oxford: Oxford University Press.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press.

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

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

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.

Richter S, Wenzel A, Stein M, Gabdoulline RR, Wade RC. 2008. webPIPSA: a web server for the comparison of protein interaction properties. *Nucleic Acids Res*. 36:W276–W280.

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.

Rooney AP, Zhang J, Nei M. 2000. An unusual form of purifying selection in a sperm protein. *Mol Biol Evol*. 17:278–283.

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

Sawyer SA. 1999. GENECONV: a computer package for the statistical detection of gene conversion. St. Louis (MO): Department of Mathematics Washington University.

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

Strahl B, Allis CD. 2000. The language of covalent histone modifications. *Nature* 403:41–45.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596–1599.

Tanaka M, Hennebold JD, MacFarlane J, Adashi EY. 2001. A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for oocyte-specific cleavage stage histones (cs-H1) of sea urchin and the B4/H1M histone of the frog. *Development*. 128:655–664.

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

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.

Ueda K, Kinoshita Y, Xu ZJ, Ide N, Ono M, Akahori Y, Tanaka I, Inoue M. 2000. Unusual core histones specifically expressed in male gametic cells of *Lilium longiforum*. *Chromosoma* 108:491–500.

Ueda K, Tanaka I. 1995. The apperance of male gamete-specific histone gH2B and gH3 during pollen development in *Lillium longiflorum*. *Dev Biol*. 169:210–217.

van Holde KE. 1988. Chromatin. New York: Springer-Verlag.

Wade RC, Gabdoulline RR, De Rienzo F. 2001. Protein interaction property similarity analysis. *Intl J Quantum Chem*. 83:122–127.

Wolfe SA, Anderson JV, Grimes SR, Stein GS, Stein JL. 1989. Comparison of the structural organization and expression of germinal and somatic rat histone H4 genes. *Biochim Biophys Acta*. 1007:140–150.

Wolfe SA, Grimes SR. 1991. Protein-DNA interactions within the rat histone H4t promoter. *J Biol Chem*. 266:6637–6643.

Wright F. 1990. The 'effective number of codons' used in a gene. *Gene*. 87:23–29.

Zalensky AO, Siino JS, Gineitis AA, Zalenskaya IA, Tomilin NV, Yau P, Bradbury EM. 2002. Human testis/sperm-specific histone H2B (hTSH2B). Molecular cloning and characterization. *J Biol Chem*. 277:43474–43480.

Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A*. 95:3708–3713.