Chapter 8
Long-Term Evolution of Histone Families: Old Notions and New Insights into Their Mechanisms of Diversification Across Eukaryotes

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

Abstract In eukaryotes and some archaebacteria, DNA is found associated with histones in a nucleoprotein complex called chromatin, which allows for a high extent of compaction of genomic DNA within the limited space of the nucleus. Early studies led to the notion that histones exhibit a conserved structural gene organization and limited protein diversity. However, research has been progressively accumulating to demonstrate that the structure, configuration and copy number of histone genes varies widely across organisms as a result of a long-term evolutionary process that promotes genetic variation. This genetic diversity is mirrored by the structural and functional diversity exhibited by the protein members of the different histone families that is, in most instances, concomitant with the complexity of the organism. The present chapter is aimed at providing a comprehensive review of the most recent information on the origin of eukaryotic histone multigene families. Particular attention is paid to the structural and functional constraints acting on histones and their relevance for the progressive diversification of histone variants during evolution, especially as it pertains to histone gene organization and expression.

R. González-Romero and J. Méndez
Departamento de Biología Celular y Molecular, Universidade da Coruña, E15071, A Coruña, Spain
D. Dryhurst and J.Ausió
Department of Biochemistry and Microbiology, University of Victoria, V8W 3P6 Victoria, Canada
J.M. Eirín-López
Departamento de Biología Celular y Molecular, Facultade de ciencias, Universidade da Coruña, Campus de A Zapateria s/n, E15071 A Coruña, Spain
e-mail: jeirin@udc.es

P. Pontarotti (ed.), *Evolutionary Biology: Concept, Modeling, and Application*,
DOI: 10.1007/978-3-642-00952-5_8,

8.1 Introduction

In eukaryotes and some archaebacteria, DNA is found associated with histones in a nucleoprotein complex called chromatin. Chromatin allows for a high extent of compaction of genomic DNA within the limited space of the nucleus and also provides the support on which most DNA metabolic functions (i.e., replication, transcription and repair) take place. The repetitive subunit of chromatin, the nucleosome, consists of an octamer of core histones (two of each H2A, H2B, H3 and H4) around which two left handed superhelical turns of DNA are wrapped (van Holde 1988). The nucleosome core particles (NCPs) are joined together in the chromatin fiber by short stretches of linker DNA that, interact with linker H1 histones, resulting in an additional folding of the chromatin fiber. Although all the domains in eukaryotic chromatin share a common nucleosomal structure, the dynamic processes responsible for the local heterogeneity observed across the genome are regulated by three principal mechanisms: the replacement of canonical histones with specialized variants that have dedicated functions (Malik and Henikoff 2003), the occurrence of histone posttranslational modifications (Jenuwein and Allis 2001) and the association with remodeling complexes responsible for nucleosome mobilization (Owen-Hughes 2003). The wide range of possible configurations that facilitate different chromatin metabolic processes are the result of the synergistic action of the aforementioned mechanisms.

Early studies on histone genes led to the formulation of several general dogmas about their structure: (a) histone genes were considered to be intronless, encoding highly conserved proteins; (b) they existed in multiple copies closely clustered in the genome and organized into tandemly repeated blocks; (c) except for the case of relatively few replacement histone variants, the mRNAs transcribed from the genes of the canonical histones were not polyadenylated and their expression was largely coupled to S phase of the cell cycle; and (d) the evolution of the histone gene families was thought to be subject to a process of concerted evolution through rapid interlocus recombination or gene conversion (Hentschel and Birnstiel 1981; Kedes 1979; Maxson et al. 1983a; Ohta 1983). However, research has been progressively accumulating to demonstrate that the organization and copy number of eukaryotic histone genes varies widely across organisms. Additional evidence that argues against the assumptions made by these dogmas was found in the case of specialized histone variants. These proteins are encoded by intron-containing genes and are expressed constantly at basal levels throughout the cell cycle (Ausió 2006). These diverse patterns of organization and expression differ even among closely related organisms and most likely reflect the presence of different regulatory mechanisms and a complex mode of gene family evolution (Doenecke et al. 1997; Khochbin 2001; Wells and Kedes 1985). The high degree of structural and functional diversification observed across different histone types provides a compelling argument against the notion of concerted evolution (homogenization) as the major force guiding long-term histone evolution.

The present chapter is aimed at providing a comprehensive review of the most recent information on the origin of eukaryotic histone multigene families and the subsequent mechanisms guiding their long-term evolution. Special attention is paid to the structural and functional constraints acting on histone proteins and their relevance to the progressive diversification of histone variants during evolution, especially as they pertain to histone gene organization and expression. Finally, a model summarizing the process of histone diversification and differentiation from their archael origin to the wide diversity of specialized variants in eukaryotes is presented.

8.2 Histone Genes Display Highly Heterogeneous Organization Patterns Across Eukaryotic Genomes

Histone gene organization has been studied for more than 30 years (Hentschel and Birnstiel 1981; Isenberg 1979; Kedes 1979; Maxson et al. 1983a) and during most of this early period, the concept of a homogeneous gene organization was believed to be the normal arrangement in the genomes of most model organisms studied at the time, especially *Drosophila* and sea urchin (Lifton et al. 1977; Maxson et al. 1983b). However, with the progressive characterization of histone genes in a broader range of organisms, a more complex picture started to emerge showing a dispersal and diversification of the genes that was apparently concomitant with the position of the different organisms on the phylogenetic tree (Doenecke et al. 1997). A great diversity of histone gene organization patterns was observed, however, it was not until histone variants were first identified and their specific functions progressively deciphered that such an extensive heterogeneity was definitively determined to be an intrinsic feature of the different histone families. In sharp contrast to canonical histones, the histone variants have a unique genomic organization (solitary genes), copy number (single-copy), gene structure (presence of introns) and regulation (basal continuous expression throughout the cell cycle) (Ausió 2006). Clearly there is no doubt that the diversification process experienced by histone genes has allowed for the progressive differentiation of histone variants with dedicated functions and this has shaped the complex, efficient and tightly controlled mechanisms of DNA packaging and regulation of chromatin dynamics in the cell nucleus. In order to illustrate the relevance of such diversity, we will next summarize the major modes of histone gene organization across representative groups of eukaryotes.

8.2.1 Prokaryotic Chromatin and the Origin of Histones

Most studies on prokaryotic chromatin have initially focused on the eubacteria *Escherichia coli*, which has a 4.6 Mb circular chromosome organized into

supercoiled domains (Worcel and Burgi 1972). The HU proteins are the most abundant proteins in the nucleoid of this organism (Murphy and Zimmerman 1997) and DNA compaction is achieved as a result of the exogenous pressure of macromolecular crowding from the nucleoplasm and from supercoiling introduced by architectural proteins and topoisomerase activity (Sandman et al. 1998). The absence of a nucleosome-based organization of bacterial chromatin contrasts with that observed in archaebateria, which contain histones as well as other chromatin associated proteins (Grayling et al. 1996; Sandman and Reeve 2005). The phylum *Euryarchaea* (Fig. 8.1) is characteristic among *Archaea* in having histone-like proteins that are relatively small in size, do not have N- or C-terminal unstructured tails comprising only a histone-fold domain (Arents and Moudrianakis 1995) and appear to form analogous or perhaps homologous structures to those formed by the eukaryotic $(H3 + H4)_2$ histone tetramers (Pereira et al. 1997). These proteins bind and wrap DNA into nucleosomal-like structures protecting about 60 bp of DNA and can induce DNA supercoiling (Sandman et al. 1998). Histone-encoding genes have also been identified in marine *Crenarchaea*, giving strong support for the argument that histones evolved very early, after the divergence of *Bacteria-Archaea* but before the separation of the archaeal and eukaryotic lineages (Cubonova et al. 2005). The ancestry shared by archaeal and eukaryotic core histones is indeed

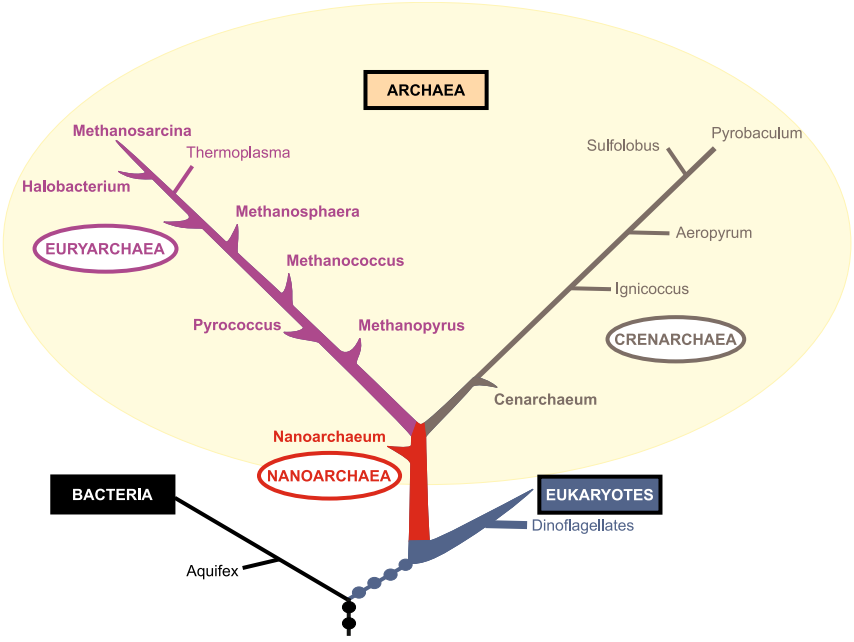

Fig. 8.1 Universal tree of life based on rRNA sequences of selected Archaea (Sandman and Reeve 2006). Different domains as well as archaeal phyla are indicated in the diagram. Thick fancy lines account for those lineages in which histone-fold sequences have been identified while dashed lines indicate uncertainty regarding the time of histone-fold origin

manifested in their conserved histone-fold region as well as in the amino acid residues that are important for histone–DNA and histone dimer–dimer interactions. Furthermore, archaeal histone sequences exhibit a degree of variation that provides clues to the molecular requirements of the evolutionary transition from prokaryotic to eukaryotic chromatin (Sandman and Reeve 2006; Zlatanova 1997).

8.2.2 *The Transition Toward the Eukaryotic Cell and the Appearance of Pluricellularity in Light of Histone Diversification*

Histones have been extensively characterized in eukaryotic genomes where, with the only exception of some protozoa such as dinoflagellates (Herzog and Soyer 1981), they organize chromatin into a repetitive nucleosome structure. Among lower eukaryotes, yeast is unique in its small number of histone genes, having only two gene copies of each of the four core histones. The H2A and H2B genes are adjacent to one another, they are divergently transcribed and they exist as two genetically unlinked copies encoding two structural variants. This is in contrast to the H3 and H4 genes that encode identical products (van Holde 1988). There is also a candidate yeast linker histone (Hho1p) encoded by a gene that is co-expressed with the core histone genes in S phase (Landsman 1996). This protein has an unusual tertiary structure with two regions (GI and GII) homologous to the single globular domain of the linker histone of higher eukaryotes (Ali and Thomas 2004). The organization of histone genes is also known in other lower eukaryotes such as the ciliates *Stylonychia lemnae*, *Tetrahymena thermophila* and closely related species, where the histone genes are unclustered in the genome and the H1 genes are apparently absent in the micronucleus (Allis et al. 1979; Prescott 1994). Also, single-copy H1 genes independently organized from core histone genes were described in *Volvox* (Lindauer et al. 1993).

Despite the classical notion of histone genes being clustered and tandemly repeated, many arrangements are observed in metazoan genomes as indicated in Fig. 8.2. For instance, in the cnidarian *Acropora formosa* (coral) histone genes consist in clusters of the four core histones (quartets) that are reiterated about 150 times (Miller et al. 1993). A similar organization has also been described in the nematode *Caenorhabditis elegans* that has independently organized single-copy H1 genes (Sanicola et al. 1990). Similarly, the histone genes of annelids are organized into clusters containing H1 and the four core histone genes (quintets) but quartets of core histones that are reiterated about 600–650 times in the genome are also present (del Gaudio et al. 1998; Sellos et al. 1990). However, molluscs represent the paradigm of diversity among metazoans showing up to three different arrangements of histone genes even within a single species. This is the case with the mussel *Mytilus galloprovincialis* (Eirín-López et al. 2004b), which displays clustered quintets linked to 5S rRNA genes, quartets of core histones and independent clusters containing replication independent H1 genes

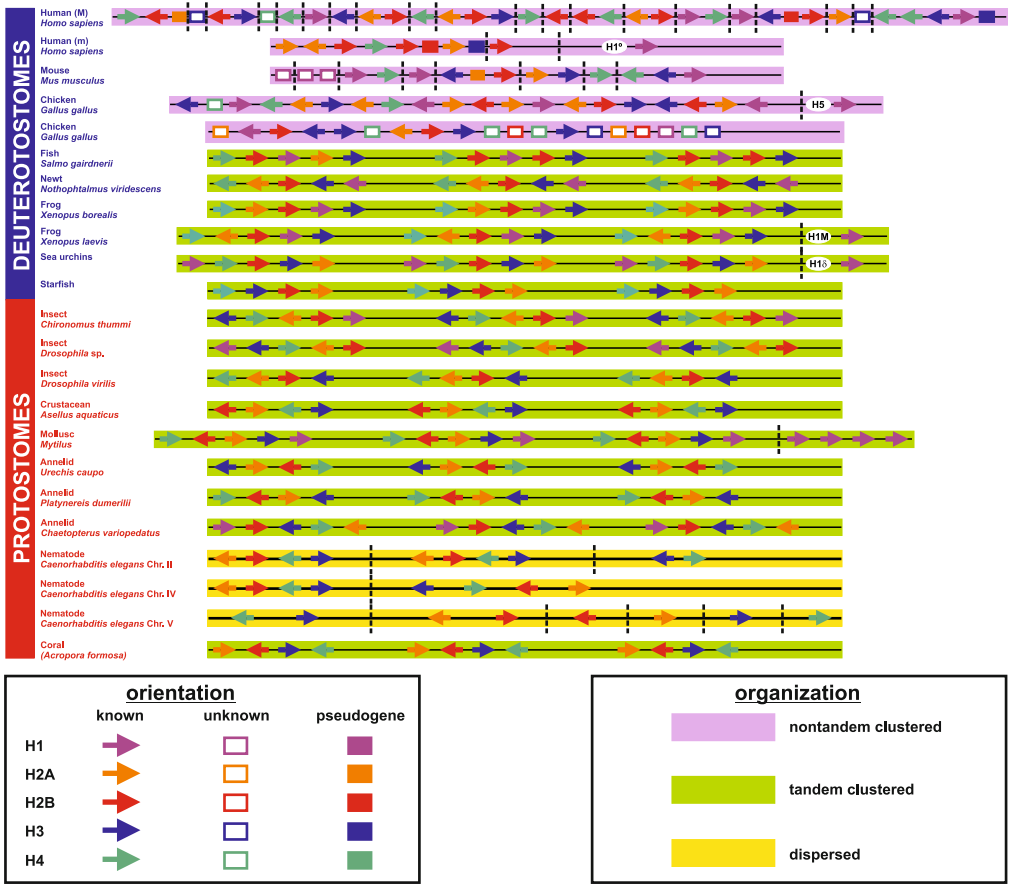

Fig. 8.2 Genomic organization of genes encoding histones across different metazoan lineages highlighting the many possible arrangements in nontandemly clustered, dispersed and tandemly clustered repeats. Directions of transcription are indicated when known, as well as the presence of pseudogenes for the five histone families and some of the replication-independent H1 variants. Both the major (M) and minor (m) human histone clusters are represented, located on chromosomes 6 and 1, respectively. Although two clusters also exist in mice, only a portion of the major cluster is shown. Two histone clusters are also present in the chicken genome but it is not certain whether these are on the same chromosome or not

(Drabent et al. 1999; Eirín-López et al. 2002). In insects, the clustered quintets containing H1 genes are present in *Drosophila* (Kremer and Henning 1990; Lifton et al. 1977; Tsunemoto and Matsuo 2001) and in the midge *Chironomus* (Hankeln and Schmidt 1991). However, clustered quartets of core histones and independent H1 histone genes are present in *Drosophila virilis* (Domier et al. 1986; Schienman et al. 1998) and *Chironomus thummi* (Hankeln and Schmidt 1993).

Echinoderms are probably one of the groups that have been more extensively analyzed among deuterostomes. The sea urchins *Strongylocentrotus purpuratus* and *Psammechinus miliaris* display around 700 repetitions of quintets tandemly arranged and expressed at early embryonic stages (Sures et al. 1978), which coexist in the genome with additional clusters without any regular organization that are expressed later on during development (Maxson et al. 1983b). However, only clustered quartets with variable arrangements have been reported in the starfish species *Pisaster ochraceus*, *P. brevispinus* and *Dermasterias imbricata* (Cool et al. 1988). Although histone genes still remain organized in clusters in many vertebrate genomes, such regular organization appears to be gradually lost during the course of evolution. For instance, amphibians display histone genes organized either in clustered quartets (newts) or quintets (*Xenopus*) with variable copy number (Ruberti et al. 1982; Stephenson et al. 1981; Turner and Woodland 1983; Turner et al. 1983; van Dongen et al. 1981), as do fishes such as *Salmo gairdnerii*, which have quintets repeated about 150 times in their genome (Connor et al. 1984). Although chickens also contain clustered histone quintets repeated about 10 times, they also have solitary and replication independent H1-type genes encoding the specialized H5 histone typical of nucleated erythrocytes (D'Andrea et al. 1985; Scott and Wells 1976). The tandem arrangement of histone genes is definitively lost in mammals. Histones are organized in either a single cluster (located on chromosome 17 in the rat [Walter et al. 1996]) or in two physically independent major and minor clusters on chromosomes 6 and 3 and chromosomes 13 and 3 in humans and mice, respectively (Albig et al. 1997a, b; Wang et al. 1997). In addition, solitary single-copy replication independent H1-type genes are also present in both human and murine genomes, encoding histone $H1^0$ that has dedicated functions in terminally differentiated cellular systems (Doenecke and Alonso 1996).

8.3 Histone Variants Impart Specific Functions to Nucleosomes

8.3.1 Linker Histones

The histone H1 family encompasses one of the largest numbers of isoforms among histones. This diversity is especially evident in mammals where 11 different linker histones have been identified: 7 somatic (H1.1–H1.5, $H1^0$ [Ausió and Abbott 2004; Parseghian and Hamkalo 2001], and H1x [Happel et al. 2005]), 3 sperm-specific (H1t [Seyedin and Kistler 1979], H1t2 [Martianov et al. 2005] and HILS1 [Iguchi et al. 2003; Yan et al. 2003]); and the oocyte-specific variant H1oo (Tanaka et al. 2001).

Other cleavage-specific H1 variants homologous to H1oo have also been described in amphibians (embryonic linker histone H1M or B4 from *Xenopus* [Cho and Wolffe 1994]) and echinoderms (histone CS from sea urchin [Mandl et al. 1997]).

The evolution of H1 has favored the differentiation of other highly specialized isoforms such as histone H5 (Ruiz-Carrillo et al. 1983), a replication-independent H1 variant restricted to the nucleated erythrocytes of birds, which also appears to be present in amphibians (Khochbin 2001) and reptiles. Therefore, H5 is structurally related to $H1^0$, a histone that replaces somatic H1 isoforms in terminally differentiated cells of many vertebrates (Panyim and Chalkey 1969). These variants have been extensively studied and they share characteristic conserved elements in their promoter regions, which are involved in their replication-independent pattern of expression. These include a UCE element (Upstream Conserved Element), an H1 box followed by a G/C rich segment and an H4 box (Eirín-López et al. 2005; Schulze and Schulze 1995). Although both the $H1^0$ and H5 genes encode polyadenylated mRNAs, the H5 transcript contains two additional stem-loop signals in the 3′ UTR region (Doenecke and Alonso 1996). Until recently, the occurrence of replication independent H1 variants had been restricted exclusively to deuterostomes. Although different studies postulated the existence of H1 histone variants in several protostomes (Ausió 1999; Barzotti et al. 2000; del Gaudio et al. 1998; Eirín-López et al. 2002, 2004b; Hankeln and Schmidt 1993) the presence of replication-independent H1 forms has only been studied in detail in molluscs (Eirín-López et al. 2005).

8.3.2 Core Histones

Since histones must be synthesized in stoichiometric amounts for the assembly of chromatin onto newly replicated DNA, transcription of histone genes is tightly regulated during the cell cycle and the bulk of their translation is coordinated with DNA replication during S phase (Marzluff 1992). A unique feature of these replication-dependent histone mRNAs is their lack of polyadenylated tails (replaced by a stem-loop signal) and the regulation of their expression at three different levels including transcriptional, mRNA processing and mRNA stability (Doenecke et al. 1997). Although these regulatory mechanisms are characteristic of most histone genes, there is a small fraction of histones encoded by solitary single-copy genes whose expression prevails in nonproliferating cells, the so-called replacement histones or histone variants (Henikoff and Ahmad 2005; Smith et al. 1984). Histone variant genes exhibit constant basal replication-independent expression throughout the cell cycle, their mRNAs contain long 3′ UTR regions as well as polyadenylated tails that bind the poly(A) binding protein that increases their stability (Marzluff 1992).

Although core histones are far more conserved than H1 histones (Isenberg 1979), this does not preclude the existence of a marked functional differentiation among their variants. This is particularly noticeable in the case of the H2A family

(Ausió and Abbott 2002), which includes the heavily studied variants H2A.X and H2A.Z. These two variants are involved in the maintenance of genome integrity and in the regulation of chromatin dynamics. Histone H2A.X is present in almost all eukaryotes and is particularly expressed in germinal cells where it has functions related to DNA repair and chromosome condensation (Li et al. 2005). This variant is encoded by intron-less genes that are expressed as two different types of transcripts: a short replication-dependent type with a stem-loop signal and a longer replication-independent type that is polyadenylated (Alvelo-Ceron et al. 2000). Histone H2A.Z has been ascribed multiple functions that may differ among species. A growing body of evidence suggests that it participates in regulation of gene expression (Barski et al. 2007; Bruce et al. 2005) but it also plays an important role in the heterochromatin structure of the centromere (Greaves et al. 2007). In vertebrates, H2A.Z exists as a mixture of two protein forms H2A.Z-1 (previously H2A.Z) and H2A.Z-2 (previously H2A.F/Z or H2A.V) that differ by three amino acids (Coon et al. 2005). These two proteins are encoded by separate genes that contain four introns in humans and are expressed through polyadenylated mRNAs (Hatch and Bonner 1988, 1990). The H2A family also includes another highly specialized variant, macroH2A. This variant is characterized by a long non-histone C-terminal tail and has been shown to be involved in female X chromosome inactivation in mammals and birds (Costanzi and Pehrson 1998; Ellegren 2002). In contrast, histone H2A.Bbd is a highly variable quickly evolving mammalian H2A variant (Eirín-López et al. 2008), which is markedly deficient in the inactive X-chromosome and participates in the destabilization of nucleosomes and in the unfolding of the chromatin fiber (González-Romero et al. 2008b).

Other core histone variants include histone H3.3, CENPA and H3.t of the H3 family. In mammals, H3.3 differs from canonical replication-dependent H3.1 by only four amino acids and is enriched in actively transcribed regions of the genome in somatic cells (Mito et al. 2005). Two identical proteins, H3.3A and H3.3B, which are encoded by two different solitary genes, participate in the transition from histones to protamines during spermiogenesis in mammals (Henning 2003; Doenecke et al. 1997; Wells and Kedes 1985). Centromeric protein A (CENPA) is involved in the packaging of chromatin at eukaryotic centromeres (Govin et al. 2005; Palmer et al. 1987) and a testis-specific H3 histone (H3t) has been described in humans (Witt et al. 1996). With regard to the other histone families, the human H2B.1 gene represents the only replication-independent H2B isoform known and encodes transcripts alternatively processed at 3′ UTR regions, yielding replication-dependent and replication-independent mRNAs (Collart et al. 1992). Although no replacement subtypes have been described for the most highly conserved family of histones, the H4 family, an H4 protein with replication-independent properties has been described in *Drosophila* (Akhmanova et al. 1996).

In striking contrast to the animal kingdom, plant histones provide a very different example of histone gene structure and regulation (Kanazin et al. 1996). This is demonstrated by the existence of an intron containing H3 genes in soybean, barley and wheat, which were found to be expressed in different plant organs in a relatively replication-independent fashion. Plant histones do not have stem-loop

signals and are transcribed into polyadenylated mRNAs (reviewed by Chabouté et al. [1993]) suggesting that the mechanisms regulating histone expression are very different between the plant and animal kingdoms. The regulation of histone production in plants would essentially occur at the transcriptional level.

8.4 Eukaryotic Histones Arose from Archaeal Histones Following a Recurrent Gene Duplication Process

The evolutionary origin of eukaryotic histones can be traced back to prokaryotes. Histones resembling H2A and H4, as well as the DNA topoisomerase V (a prokaryotic counterpart of eukaryotic topoisomerase I) were found in the hyperthermophilic archaebacteria *Methanopyrus kandleri* (Slerasev et al. 1984) and also archaeal RNA polymerases show common features with eukaryotic RNA polymerases (Reeve et al. 1997). In addition, archaeal histones exhibit some similarities with the central domain of the CBF-A eukaryotic transcription factor subunit (CCAAT-binding factor subunit A), suggesting that the eukaryotic modes of transcription and DNA packaging may have originated before the eukaryotes themselves (Ouzounis and Kyrpides 1996). Consistent with this observation, phylogenetic trees reconstructed from rRNA data place *Archaea* and *Eukarya* on the same branch, indicating the existence of a common ancestor shared by both groups after the divergence of *Archaea* and *Eubacteria* (Sandman and Reeve 1998, 2006).

The presence of histone-like genes in *Euryarchaeota* was explained by Sandman and Reeve (1998) in terms of the 'hydrogen hypothesis for the first eukaryote' (Martin and Müller 1998). Accordingly, the origin of the eukaryotic cell was proposed to have been derived from a symbiotic association between a methanogen archaebacterium and a proteobacterium. Thus, while the eukaryotic nucleus would have been derived from the archaebacterium (including the proteins involved in DNA packaging), most of the cellular metabolism would have been contributed by the proteobacterium. Consequently, the evolution of histones and DNA condensation into nucleosomes apparently occurred in the euryarchaeotal lineage, before the archaeal-eukaryal divergence, facilitating the genome expansion and the development of *Eukarya* (Sandman and Reeve 2005, 2006).

It appears that DNA duplication has been the basis underlying the evolutionary mechanisms driving histone evolution from the early appearance of these proteins in *Archaea* and the evolutionary process shaping this variability can be summarized in three major stages as indicated in Fig. 8.3. Firstly, the evolution of histones in *Archaea* would have been based on their generic role in packaging the genomic DNA. In this context, the first important event of gene duplication would have resulted in at least two histone-like genes in *Archaea* (such as HMfA and HMfB genes identified in *Methanothermus fervidus*, Fig. 8.3a) capable of forming tetrameric complexes to efficiently compact DNA (Sandman et al. 1990; Starich et al. 1996). The next critical duplication event would be represented by the intragenic

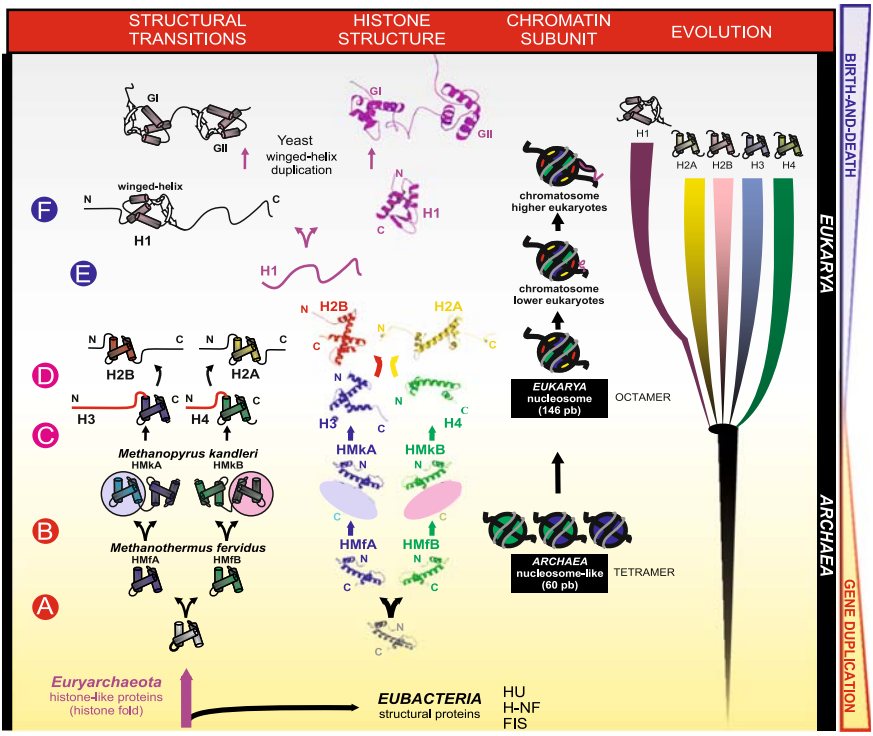

Fig. 8.3 Schematic representation of histone evolution since their origin in archaebacteria through recurrent gene duplication steps, which represent the underlying mechanisms driving histone evolution and the selective mechanisms shaping histone variability. Transitions in histone gene structure were immediately mirrored by the configuration of the nucleosome core particle (NCP) regarding its ability to package DNA. The progressive evolutionary diversification and specialization of core histones is indicated in the right margin, concomitantly with the incorporation of linker histones into the chromatin structure

duplication giving rise to the doublet histones (Fig. 8.3b) such as those identified in *M. kandleri* (Malik and Henikoff 2003). In this scenario it would be plausible that while one of the two histone folds retained its critical role in DNA packaging, the additional one would have been available to provide a substrate for selection experiments, eventually resulting in the further functionalization of the histone N- and C- terminal tails (Fig. 8.3c).

In the second stage, the histone doublet structure resulted in the formation of an asymmetric dimer that would have been the direct predecessor of the cannonical H3-H4 dimers. The evolutionary constraints generated in the transition towards a eukaryotic-specific mitosis would have been strong enough to force the shift towards an octamer structure by incorporating additional H2A–H2B dimers (Fig. 8.3d), a process that may have facilitated the genome expansion and the development of *Eukarya* (Malik and Henikoff 2003). Linker histones were

likely the last component to join this structure, providing the maximum level of compaction of DNA. However, it is still not clear how this process took place. The C-terminal tail of eukaryotic H1 histones is able to compact chromatin by itself and this domain constitutes the whole protein in the case of early ancestral eukaryotes such as trypanosomes (Grüter and Betschart 2001). This supports the hypothesis that, in contrast to core histones, ancestral H1 histones were composed of only a C-terminal region (Fig. 8.3e) and the core domain containing the winged-helix motif would have been acquired later in evolution as indicated in Fig. 8.3f (Kasinsky et al. 2001).

In the third stage, the differentiation of the five metazoan histones (H1, H2A, H2B, H3 and H4) would have marked the beginning of the final stage in histone evolution. Although gene duplication has prevailed as the major mechanism in providing the eukaryotic cell with the required amounts and diversity of histones (Malik and Henikoff 2003), it has been shown that concerted evolution does not play a major role in their evolution (Eirín-López et al. 2004a; Eirín-López et al. 2005; Piontkivska et al. 2002; Rooney et al. 2002). Crucial to this stage is the diversification as a result of recurrent gene duplications followed by a strong purifying selection process acting at the protein level. This process is known as birth-and-death evolution (Nei and Hughes 1992; Nei et al. 1997) and represents the major mode of long-term evolution in eukaryote histone families as well as in many other gene families (Nei and Rooney 2006), as discussed in detail in the following section. The evolutionary refinement resulting from such diversification would have been determined by the cellular specialization associated with the appearance of multicellular organisms. Histone variability in these organisms is required in order to accommodate the different packing needs and regulation of gene expression in different cell types and developmental stages.

8.5 The Long-Term Evolution of Histone Genes Is Guided by a Birth-and-Death Process That Promotes Genetic Diversity

In contrast to the notion of divergent evolution, until around 1990 most multigene families were thought to be subject to concerted evolution (Fig. 8.4). Concerted evolution involves a process where a mutation occurring in a repeat spreads all through the gene family members by repeated occurrence of unequal crossover or gene conversion (Arnheim 1983; Smith 1974). The validity of this model, in the case of histones, was further reinforced by the general view that a gene family that produces a large amount of gene products is subject to concerted evolution in order to maintain the homogeneity of the protein product (Baldo et al. 1999; Coen et al. 1982; Kedes 1979; Liao 1999; Matsuo and Yamazaki 1989; Thatcher and Gorovsky 1994). However, as more amino acid and DNA sequence information became available, some serious conceptual concerns arose when trying to apply the

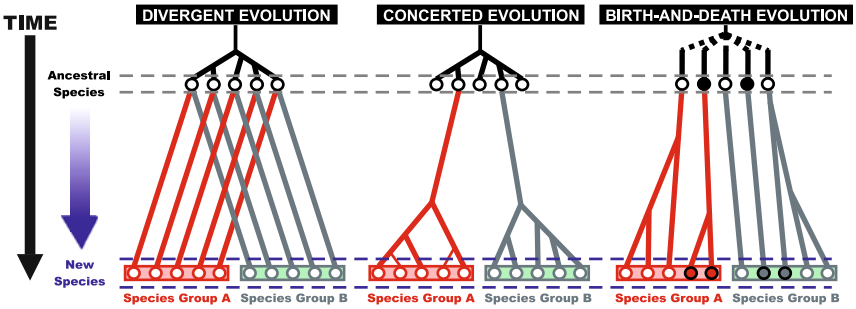

Fig. 8.4 Schematic representation of the three major models of multigene family evolution proposed for histone gene families since their discovery (Nei and Rooney 2006). Divergent evolution was the first mechanism proposed in order to explain the long-term evolution of hemoglobin α, β, γ and δ chains and myoglobin, whose encoding genes have diverged gradually as the duplicate genes acquired new functions. However, the close intraspecific relationship among ribosomal RNA genes was very difficult to reconcile with the aforementioned mechanism, leading to an idea based on a process of unequal crossover or gene conversion responsible for the homogenization of the family members, known as concerted evolution. Later on, the availability of DNA and protein sequences allowed to revisit the mechanisms guiding the evolution of some gene families such as histones, questioning the applicability of concerted evolution and suggesting a model called birth-and-death, based on recurrent gene duplications (open and solid circles indicate active genes and pseudogenes in this model, respectively)

concepts of concerted evolution to some gene families. These include genes involved in immune systems and disease-resistance (Hughes and Nei 1990; Ota and Nei 1994; Zhang et al. 2000), as well as highly conserved gene families such as ubiquitins and histones (Eirín-López et al. 2004a, 2005; González-Romero et al. 2008a; Nei et al. 2000; Piontkivska et al. 2002; Rooney et al. 2002).

As mentioned in the previous section, the differentiation of the five eukaryotic histone families, together with the subsequent specialization of the histone variants, represented an evolutionary breakthrough that allowed for a maximal level of chromatin compaction and structural and functional diversification. The broad gene and protein diversity of histones is seemingly contradictory to what would be expected from a concerted evolution model since such a model would predict close intraspecific relationships among histone genes. Several studies have dealt with this paradigm during recent years by analyzing protein and nucleotide variation levels within the different histone gene families across different groupings of eukaryotes. However, none of these studies found any support for a process of homogenization being involved in histone evolution and there are at least three major lines of evidence that argue against such a process. Firstly, phylogenetic inference of the evolutionary history of histone genes revealed that different histone isoforms cluster by type instead of by species in the phylogenetic trees. Furthermore, the analyses of nucleotide sequences failed to detect any signal of homogenization among histone genes. For instance, comparisons between human and mouse H1 genes showed that paralogous genes are not more closely related than orthologous H1 genes from both species, indicating that the functional

differentiation of these genes is most likely due to a process involving selection rather than homogenization (Eirín-López et al. 2004a). A second line of evidence arose from the analyses of the nature of nucleotide variation occurring among histone genes, which indicated that the synonymous nucleotide divergence was always significantly larger than the nonsynonymous variation (Eirín-López et al. 2004a; González-Romero et al. 2008a; Piontkivska et al. 2002; Rooney et al. 2002). This suggests the absence of a concerted evolution process and is again consistent with a selective process acting at the protein level. Finally, pseudogene evolution provides a powerful tool for examining the presence of homogenization across histone repeats, given that they are expected to have a lower level of divergence compared to active genes. In this regard, all the studies focused on the long-term evolution of histone families have detected significant levels of divergence in pseudogenes (Eirín-López et al. 2004a; González-Romero et al. 2008a; Piontkivska

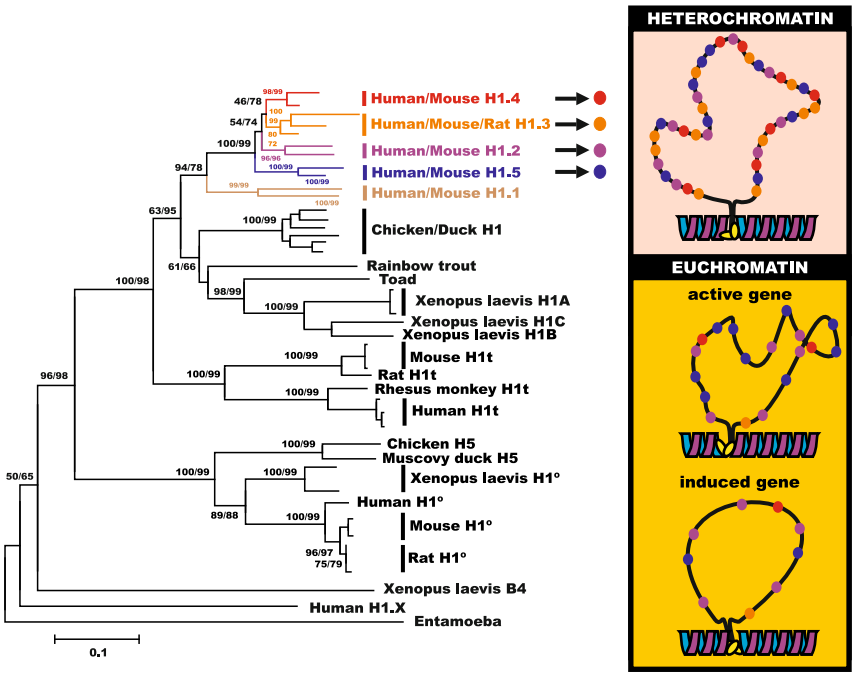

Fig. 8.5 Phylogenetic relationships among protein sequences of members of the H1 family in chordates showing a functional clustering pattern in which orthologs are more closely related among them than to their respective paralogs (Eirín-López et al. 2004a). Although histone H1 somatic types from mammals (H1.1–H1.5) were thought to be functionally redundant, the topology depicts a process of functional differentiation leading to different H1 identities. This notion has been indeed supported by experiments revealing a nonrandom distribution of H1 subtypes in the chromatin fiber, related to different transcriptional states (Parseghian et al. 2000). Numbers in internal nodes indicate confidence intervals calculated as bootstrap and internal brach-test, respectively

et al. 2002; Rooney et al. 2002), discarding any significant effect resulting from concerted evolution.

All the above mentioned studies proposed that the results fit better with the birth-and-death model of evolution (Nei and Hughes 1992) in which the diversification of the family members is the result of a recurrent process of gene duplication (mutation) followed by the inactivaction of some of the newly generated genes (selection, see Fig. 8.4 for details). Homogeneity is maintained by the effect of a strong purifying selection at the protein level and thus, DNA sequences of different gene family members can be very divergent both within and between species (Nei and Hughes 1992; Nei et al. 1997). A good example of this mode of long-term evolution is best illustrated by the histone H1 family, where extensive knockout experiments have demonstrated the existence of a large amount of functional redundancy among somatic replication-dependent H1 subtypes. However, there is now evidence indicating that the four histone H1 subtypes that are present in all mammalian somatic cells are not randomly distributed in chromatin (Fig. 8.5). This provides evidence for different roles for the different somatic H1 subtypes, in agreement with the evolutionary picture of the functional differentiation of H1 histones (Eirín-López et al. 2004a; Parseghian et al. 2000).

The birth-and-death model of evolution, as opposed to the gene homogenization process that would result from concerted evolution, promotes genetic variation. Therefore, it provides quite a reasonable mechanism for explaining the long-term evolution of gene families with high levels of diversity among their members, such as histones. This does by no means imply that gene conversion or unequal crossover does not occur but it strongly suggests that their contribution to the diversification of multigene families is relatively minor.

8.6 Replication-Dependent Histone Variants Are Derived from a Common Replication- Independent Ancestor

Different histone variants are expressed in a tissue- and developmental stage-dependent manner. During the course of evolution, the origin of these variants could have arisen by two different ways: by a gradual process from an ancient differentiation event, or through multiple independent events. An ‘orphon’ origin was initially proposed to explain the evolutionary origin of replication-independent histone variants followed by a process of concerted evolution. The isolation of these genes from the main histone repeats of replication-dependent variants would account for the divergent, solitary, single-copy nature of their genes (Drabent et al. 1999; Eirín-López et al. 2002, 2004b; Schulze and Schulze 1995). The discovery that long-term histone evolution occurs by a birth-and-death mechanism (Eirín-López et al. 2004a, 2005; González-Romero et al. 2008a; Piontkivska et al. 2002; Rooney et al. 2002) forced a revision of the ‘orphon’ hypothesis for the origin of the replication-independent histone variants. The process of purifying selection acting at the protein level would have preceded the split between protostomes and

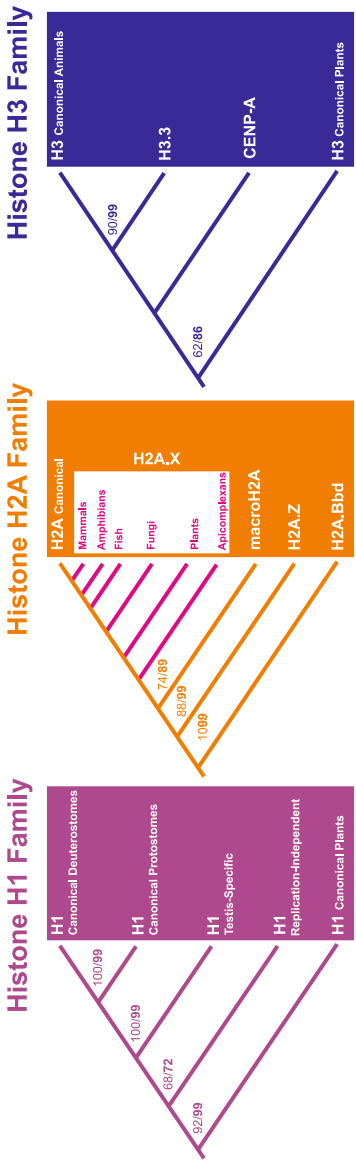

Fig. 8.6 Evolution of the major histone variant lineages within the histone H1, H2A and H3 families depicting an ancient origin except for the case of histone H2A.X. Numbers in internal nodes are indicated as in the case of Fig. 8.5

deuterostomes allowing for the subsequent transposition of replication-independent histone variants to solitary locations in the genome, where they would gradually continue their evolution (Eirín-López et al. 2004a, 2005).

Different histone families seem to be subject to different rates of 'birth-and-death' evolution as indicated by the high levels of diversity exhibited by members of the H1 and H2A families in contrast to the H2B, H3 and H4 families. Although most replication-independent histone variants have an ancient origin, there are notable exceptions to this rule, such as the case of histone H2A.X (Fig. 8.6). It is very likely that H2A.X genes arose separately and recurrently during evolution having totally replaced canonical H2A in organisms such as *Saccharomyces cerevisiae* while being completely absent in *C. elegans* (Thatcher and Gorovsky 1994). Less is known about the H3 genes, where a replication independent gene was initially proposed to be the progenitor of all H3 genes through a single differentiation event that took place early in evolution (Wells et al. 1986). This hypothesis is supported by a study that suggests that a gene similar to that of histone H3 from the protist *Phreatamoeba*, the closest relative to animal and plant H3 genes, may have been the ancestor of the animal, plant and fungal H3 sequences. Nevertheless, the appearance of H3 variants independently in animals, plants and *Tetrahymena* was also taken as evidence for the multiple origin of H3 variants (Thatcher and Gorovsky 1994).

When considering long-term histone evolution it is important to bear in mind the large differences exhibited by the plant and animal kingdoms, which clearly reflect the different evolutionary strategies followed by different organisms despite having all gone through the same histone gene duplication and selection mechanisms. Plant histones show very unique features that clearly differentiate them from the replication-dependent histones from animals but they are very closely related to animal replication-independent histone variants with which they share common traits such as the presence of introns and expression through polyadenylated transcripts (Chabouté et al. 1993). The most plausible explanation to account for this relies on the fact that plant cells exhibit a much longer cell cycle than animal cells. Therefore, a rapid change in the levels of histone gene expression during S-phase is no longer needed and transcription control has a predominant role over posttranscriptional regulation. All this raises the question of whether ancestral histones were expressed through polyadenylated transcripts. Different authors have suggested that the major plant histone genes evolved from a common polyadenylated ancestor prior to the differentiation between plants and animals (Chabouté et al. 1993) and that animal histone genes would have acquired specific posttranslational regulatory mechanisms (necessary to ensure the rapid histone biosynthesis in rapidly dividing cells) later on during evolution. This hypothesis is further supported by the polyadenylated nature of different histone transcripts in ancestral eukaryotes preceding the differentiation of the metazoan variants. A good example of this is provided by the histone H1 from trypanosomes (Grüter and Betschart 2001).

The mechanisms of transcriptional regulation of histone genes play a critical role in the specialization of histone variants in different tissues and developmental stages. The bulk of histone mRNA translation is coordinated with DNA replication during S phase of the cell cycle and this process is mediated by the presence of

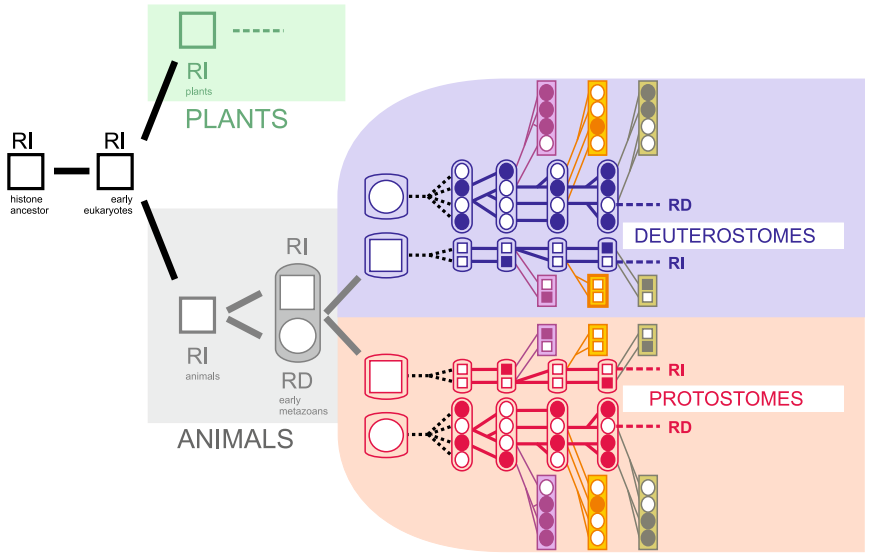

Fig. 8.7 Representation of the evolutionary process followed by an ancestral replication-independent histone leading to an initial differentiation between plant and animal histones (replication independent), which was subsequently followed by the segregation of the two animal histone lineages (replication independent, RI; replication dependent, RD) early in metazoan evolution. Both RI and RD lineages followed a parallel process of birth-and-death evolution across protostomes and deuterostomes leading to the differentiation of the canonical histones (RD) and histone variants (RI) in different taxonomic groups

a stem-loop signal in the 3′ UTR region of the transcript that is unique to replication-dependent histones in animals. In contrast, the expression of all other histone types is mediated by a polyadenylation signal that confers stability to the mRNAs, as is the case for almost any other type of gene in the genome. The fact that the 'canonical' replication-dependent histones in animals represent the only examples of genes lacking polyadenylated transcripts and given the early differentiation exhibited by the replication-independent histone variants, the notion that the primordial genes of eukaryotic histones were also transcribed through polyadenylated transcripts is supported (Fig. 8.7).

8.7 Conclusions

The old idea of histones being small basic proteins that provide a scaffold for DNA in the chromatin fiber and whose genes evolve through concerted evolution is only accurate in describing histones as small and basic proteins. As discussed throughout this work, histones have an evolutionary origin that can be traced back to archaebateria. It involves a progressive diversification and differentiation of the four core histone families through a mechanism of recurrent gene duplication, which eventually facilitated DNA compaction in the transition towards the eukaryotic cell.

The extraordinary structural and functional diversity observed among members of the different histone families is, in most instances, concomitant with the complexity of the organism, owing to the critical role histones play in the evolution of biological systems. Such diversity provides compelling evidence in support of a mechanism directing the long-term evolution of histone families that is geared towards the generation of genetic diversity (birth-and-death), rather than one that induces gene homogenization (concerted evolution). Two major histone lineages must have already been differentiated very early in eukaryotic evolution, one leading to the canonical replication-dependent histones and the other to the replication-independent histone variants. Both lineages appear to share a common replication-independent ancestor containing introns, which most likely was the preferred histone choice in plants. It is important to bear in mind that the loss of introns and the appearance of a replication-dependent expression must have had critical consequences for the evolutionary constraints acting upon the proteins. It probably involved a dramatic temporal switch towards strong positive selection of the replication-dependent histone genes that would result in an oddly similar signature to that arising from purifying selection. The current availability of genome sequences for a broad range of eukaryotic organisms opens a door to further examine the implications of such an intriguing phenomenon, especially as it pertains to the distribution and evolution of introns in early eukaryotes and early metazoans in relation to the selective pressures acting on histones.

Acknowledgments This work was funded in part by the Canadian Institutes of Health Research (CIHR) Grant MOP-57718 to Juan Ausió and by the Marie Curie Outgoing International Fellowship (MOIF-CT-2005–021900) within the 6th Framework Programme (European Union) and by a contract within the Isidro Parga Pondal Program (Xunta de Galicia) to José M. Eirín-López. Rodrigo González-Romero is the recipient of a fellowship from the Diputacion de A Coruña.

References

Akhmanova A, Miedema K, Henning W (1996) Identification and characterization of the Drosophila histone H4 replacement gene. FEBS Lett 388:219–222

Albig W, Meergans T, Doenecke D (1997a) Characterization of the H1.5 genes completes the set of human H1 subtype genes. Gene 184:141–148

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

Ali T, Thomas JO (2004) Distinct properties of the two putative “globular domains” of the yeast linker histone, Hho1p J Mol Biol 337:1123–1135

Allis CD, Glover CVC, Gorovsky MA (1979) Micronuclei of Tetrahymena contain two types of histone H3. Proc Natl Acad Sci USA 76:4857–4861

Alvelo-Ceron D, Niu L, Collart DG (2000) Growth regulation of human variant histone genes and acetylation of the encoded proteins. Mol Biol Rep 27:61–71

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

Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) Evolution of Genes and Proteins. Sinauer, Sunderland, MA, pp 38–61

Ausió J (1999) Histone H1 and evolution of sperm nuclear basic proteins. J Biol Chem 274:31115–31118
Ausió J (2006) Histone variants: the structure behind the function. Brief Funct Genomic Proteomic 5:228–243
Ausió J, Abbott DW (2002) The many tales of a tail: carboxyl-terminal tail heterogeneity specializes histona H2A variants for defined chromatin function. Biochemistry 41:5945–5949
Ausió J, Abbott DW (2004) The role of histone variability in chromatin stability and folding. In: Zlatanova J, Leuba SH (eds) Chromatin Structure and Dynamics: State-of-the-Art. Elsevier, New York, pp 241–290
Baldo AM, Les DH, Strausbaugh LD (1999) Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of *Sophophora*. Mol Biol Evol 16:1511–1520
Barski A, Cuddapah S, Cui K, Roth TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. Cell 129:823–837
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
Bruce K, Myers FA, Mantouvalou E, Lefevre P, Greaves I, Bonifer C, Tremethick DJ, Thorne AW, Crane-Robinson C (2005) The replacement histone H2A.Z in hyperacetylated form is a feature of active genes in chicken. Nucleic Acids Res 33:5633–5639
Chabouté ME, Chaubet N, Gigot C, Phillips G (1993) Histones and histone genes in higher plants: Structure and genomic organization. Biochimie 75:523–531
Cho H, Wolffe AP (1994) Xenopus laevis B4, an intron-containing oocyte-specific linker histone-encoding gene. Gene 143:233–238
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
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
Cool D, Banfield D, Honda BM, Smith MJ (1988) Histone genes in three sea star species: Cluster arrangement, transcriptional polarity and analysis of the flanking regions of H3 and H4 genes. J Mol Evol 27:36–44
Coon JJ, Ueberheide B, Syka JE, Dryhurst DD, Ausió J, Shabanowitz J, Hunt DF (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. Proc Natl Acad Sci USA 102:9463–9468
Costanzi C, Pehrson JR (1998) Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature 393:599–601
Cubonova L, Sandman K, Hallam SJ, DeLong EF, Reeve JN (2005) Histones in *Crenarchaea*. J Bacteriol 187:5482–5485
D'Andrea R, Coles LS, Lesnikowski C, Tabe L, Wells JRE (1985) Chromosomal organization of chicken histone genes: preferred associations and inverted duplications. Mol Cell Biol 5:3108–3115
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
Doenecke D, Alonso A (1996) Organization and expression of the developmentally regulated $H1^0$ histone gene in vertebrates. Int J Dev Biol 40:423–431
Doenecke D, Albig W, Bode C, Drabent B, Franke K, Gavenis K, Witt O (1997) Histones: genetic diversity and tissue-specific gene expression. Histochem Cell Biol 107:1–10
Domier LL, Rivard JJ, Sabatini LM, Blumenfeld M (1986) *Drosophila virilis* histone gene clusters lacking H1 coding segments. J Mol Evol 23:149–158

Drabent B, Kim J-S, Albig W, Prats E, Cornudella L, Doenecke D (1999) *Mytilus edulis* histone gene clusters containing only H1 genes. J Mol Evol 49:645–655
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, 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
Ellegren H (2002) Dosage compensation: do birds do it as well? Trends Genet 18:25–28
González-Romero R, Ausió J, Méndez J, Eirín-López JM (2008a) 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
González-Romero R, Méndez J, Ausió J, Eirín-López JM (2008b) Quickly evolving histones, nucleosome stability and chromatin folding: All about histone H2A.Bbd. Gene 413:1–7
Govin J, Caron C, Rousseaux S, Khochbin S (2005) Testis-specific histone H3 expression in somatic cells. Trends Biochem Sci 30:357–359
Grayling RA, Sandman K, Reeve JN (1996) Histones and chromatin structure in hyperthermophilic Archaea. FEMS Microbiol Rev 18:203–213
Greaves IK, Rangasamy D, Ridgway P, Tremethick DJ (2007) H2A.Z contributes to the unique 3D structure of the centromere. Proc Natl Acad Sci USA 104:525–530
Grüter E, Betschart B (2001) Isolation, characterisation and organisation of histone H1 genes in African trypanosomes. Parasitol Res 87:977–984
Hankeln T, Schmidt ER (1991) The organization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*. Chromosoma 105:25–31
Hankeln T, Schmidt ER (1993) Divergent evolution of an "orphon" histone gene cluster in *Chironomus*. J Mol Biol 234:1301–1307
Happel N, Schulze E, Doenecke D (2005) Characterization of human histone H1x. Biol Chem 386:541–551
Hatch CL, Bonner WM (1988) Sequence of cDNAs from mammalian H2A.Z, an evolutionarily diverged but highly conserved basal histone H2A isoprotein species. Nucleic Acids Res 16:1113–1124
Hatch CL, Bonner W (1990) The human histone H2A.Z gene. Sequence and regulation. J Biol Chem 265:15211–15218
Henikoff S, Ahmad K (2005) Assembly of variant histones into chromatin. Annu Rev Cell Dev Biol 21:133–153
Henning W (2003) Chromosomal proteins in the spermatogenesis of Drosophila. Chromosoma 111:489–494
Hentschel CC, Birnstiel ML (1981) The organization and expression of histone gene families. Cell 25:301–313
Herzog M, Soyer MO (1981) Distinctive features of dinoflegellate chromatin. Absence of nucleosomes in a primitive species *Prorocentrum micans*. Eur J Cell Biol 23:295–302
Hughes AL, Nei M (1990) Evolutionary relationships of class II major-histocompatibility complex genes in mammals. Mol Biol Evol 7:491–514

Iguchi N, Tanaka H, Yomogida K, Nishimune Y (2003) Isolation and characterization of a novel cDNA encoding a DNA-binding protein (Hils1) specifically expressed in testicular haploid germ cells. Int J Androl 26:354–365
Isenberg I (1979) Histones. Annu Rev Genet 48:159–191
Jenuwein T, Allis CD (2001) Translating the histone code. Science 293:1074–1080
Kanazin V, Blake T, Shoemaker RC (1996) Organization of the histone H3 genes in soybean, barley, and wheat. Mol Gen Genet 250:137–147
Kasinsky HE, Lewis JD, Dacks JB, Ausió J (2001) Origin of H1 histones. FASEB J 15:34–42
Kedes L (1979) Histone genes and histone messengers. Annu Rev Biochem 225:501–510
Khochbin S (2001) Histone H1 diversity: bridging regulatory signals to linker histone function. Gene 271:1–12
Kremer H, Henning W (1990) Isolation and characterization of a *Drosophila hydei* histone DNA repeat unit. Nucleic Acids Res 18:1573–1580
Landsman D (1996) Histone H1 in *Saccharomyces cerevisiae* – a double mystery solved. Trends Biochem Sci 21:287–288
Li A, Eirín-López JM, Ausió J (2005) H2AX: tailoring histone H2A for chromatin-dependent genomic integrity. Biochem Cell Biol 83:505–515
Liao D (1999) Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64:24–30
Lifton RP, Goldberg ML, Karp RW, Hogness DS (1977) The organization of the histone genes in *Drosophila melanogaster*: functions and evolutionary implications. Cold Spring Harbor Symp Quant Biol 42:1047–1051
Lindauer A, Müller K, Schmidt R (1993) Two histone H1-encoding genes of the green alga *Volvox carteri* with features intermediate between plant and animal genes. Gene 239:15–27
Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. Nat Struct Biol 10:882–891
Mandl B, Brandt WF, Superti-Furga G, Graninger PG, Birnstiel M, Busslinger M (1997) The five cleavage-stage (CS) histones of the sea urchin are encoded by a maternally expressed family of replacement histone genes: functional equivalence of the CS H1 and frog H1M (B4) proteins. Mol Cell Biol 17:1189–1200
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
Martin W, Müller M (1998) The hydrogen hypothesis for the first eukaryote. Nature 392:37–41
Marzluff WF (1992) Histone 3′ ends: essential and regulatory functions. Gene Express 2:93–97
Matsuo Y, Yamazaki T (1989) Nucleotide variation and divergence in the histone multigene family in *Drosophila melanogaster*. Genetics 122:87–97
Maxson R, Cohn R, Kedes L, Mohun T (1983a) Expression and organization of histone genes. Annu Rev Genet 17:239–277
Maxson R, Mohun T, Gormezano G, Childs G, Kedes L (1983b) Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin. Nature 301:120–125
Miller DJ, Harrison PL, Mahony TJ, McMillan JP, Miles A, Odorico DM, ten Lohuis MR (1993) Feature of histone gene structure and organization are common to diploblastic and tripoblastic metazoans. J Mol Evol 37:245–253
Mito Y, Henikoff JG, Henikoff S (2005) Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet 37:1090–1097
Murphy LD, Zimmerman SB (1997) Stabilization of compact spermidine nucleoids from *Escherichia coli* under crowded conditions: implications for in vivo nucleoid structure. J Struct Biol 119:336–346
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/England, pp 27–38
Nei M, Rooney AP (2006) Concerted and birth-and-death evolution in multigene families. Annu Rev Genet 39:121–152

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci USA 94:7799–7806
Nei M, 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
Ohta T (1983) On the evolution of multigene families. Theor Popul Biol 23:216–240
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
Ouzounis CA, Kyrpides NC (1996) Parallel origin of the nucleosome core and eukaryotic transcription from archaea. J Mol Evol 42:234–239
Owen-Hughes T (2003) Colworth memorial lecture. Pathways for remodelling chromatin. Biochem Soc Trans 31:893–905
Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL (1987) A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. J Cell Biol 104:805–815
Panyim S, Chalkey R (1969) High resolution acrylamide gel electrophoresis of histones. Arch Biochem Biophys 63:265–297
Parseghian MH, Hamkalo BA (2001) A compendium of the H1 family of somatic subtypes: an elusive cast of characters and their characteristics. Biochem Cell Biol 79:289–304
Parseghian MH, Newcomb RL, Winokur ST, Hamkalo BA (2000) The distribution of somatic H1 subtypes is nonrandom on active vs. inactive chromatin: distribution in human fetal fibroblasts. Chromosome Res 8:405–424
Pereira SL, Grayling RA, Lurz R, Reeve JN (1997) Archaeal nucleosomes. Proc Natl Acad Sci USA 94:12633–12637
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
Prescott DM (1994) The DNA of ciliated protozoa. Microbiol Rev 58:233–267
Reeve JN, Sandman K, Daniels CJ (1997) Archaeal histones, nucleosomes, and transcription initiation. Cell 89:999–1002
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
Ruberti I, Fragapane P, Pierandrei-Arnaldi P, Beccari E, Arnaldi F, Bozzoni I (1982) Characterization of histone genes isolated from Xenopus laevis and Xenopus tropicalis genomic libraries. Nucleic Acids Res 10:1544–1550
Ruiz-Carrillo A, Affolter M, Renaud J (1983) Genomic organization of the genes coding for the main six histones of the chicken: complete sequence of the H5 gene. J Mol Biol 170:843–859
Sandman K, Reeve JN (1998) Origin of the eukaryotic nucleus. Science 281:501–503
Sandman K, Reeve JN (2005) Archaeal chromatin proteins: different structures but common function? Curr Opin Microbiol 8:656–661
Sandman K, Reeve JN (2006) Archaeal histones and the origin of the histone fold. Curr Opin Microbiol 9:520–525
Sandman K, Krzycki JA, Dobrinski B, Lurz R, Reeve JN (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. Proc Natl Acad Sci USA 87:5788–5791
Sandman K, Pereira SL, Reeve JN (1998) Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. Cell Mol Life Sci 54:1350–1364
Sanicola M, Ward S, Childs G, Emmons SW (1990) Identification of a *Caenorhabditis elegans* histone H1 gene family. Characterization of a family member containing an intron and encoding poly(A) + mRNA. J Mol Biol 212:259–268
Schienman JE, Lozovskaya ER, Strausbaugh LD. (1998) *Drosophila virilis* has atypical kinds and arrangements of histone repeats. Chromosoma 107:529–539
Schulze E, Schulze B (1995) The vertebrate linker histones $H1^0$, H5, and H1M are descendants of invertebrate "orphon" histone H1 genes. J Mol Evol 41:833–840
Scott AC, Wells JR (1976) Reiteration frequency of the gene for tissue-specific histone H5 in the chicken genome. Nature 259:635–638

Sellos D, Krawetz SA, Dixon GH (1990) Organization and complete nucleotide sequences of the core-histone-gene cluster of the annelid *Platynereis dumerilii*. Eur J Biochem 190:21–29

Seyedin SM, Kistler WS (1979) H1 histone subfractions of mammalian testes. 1. Organ specificity in the rat. Biochemistry 18:1371–1375

Slerasev AI, Belova GI, Kozyavkin SA, Lake JA (1984) Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. Nucleic Acids Res 26:427–430

Smith BJ, Harris MR, Sigournay CM, Mayes ELV, Bustin M (1984) A survey of $H1^0$- and H5-like protein structure and distribution in lower and higher eukaryotes. Eur J Biochem 138:309–317

Smith GP (1974) Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp Quant Biol 38:507–513

Starich MR, Sandman K, Reeve JN, Summers MF (1996) NMR Structure of the HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. J Mol Biol 255:187–203

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

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

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

Tsunemoto K, Matsuo Y (2001) Molecular evolutionary analysis of a histone gene repeating unit from *Drosophila simulans*. Genes Genet Syst 76:355–361

Turner PC, Woodland HR (1983) Histone gene number and organisation in *Xenopus*: *Xenopus borealis* has a homogeneous major cluster. Nucleic Acids Res 11:971–986

Turner PC, Aldridge TC, Woodland HR, Old RW (1983) Nucleotide sequences of H1 histone genes from *Xenopus laevis*. A recently diverged pair of H1 genes and an unusual H1 pseudogene. Nucleic Acids Res 11:4093–4106

van Dongen W, de Laaf L, Zaal R, Moorman AF, Destree O (1981) The organization of the histone genes in the genome of *Xenopus laevis*. Nucleic Acids Res 25:2297–2311

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

Walter L, Klinga-Levan K, Helou K, Albig W, Drabent B, Doenecke D, Günther E, Levan G (1996) Chromosomal mapping of rat histone genes H1fv ($H1^0$), H1d, H1t, Th2a and Th2b. Cytogenet Cell Genet 75:136–139

Wang ZF, Sirotkin AM, Buchold GM, Skoultchi AI, Marzluff WF (1997) The mouse histone H1 genes: gene organization and differential regulation. J Mol Biol 271:124–138

Wells JRE, Kedes L (1985) Structure of a human histone cDNA: evidence that basally expressed histone genes have intervening sequences and encode polyadenylated mRNAs. Proc Natl Acad Sci USA 82:2834–2838

Wells JRE, Bains W, Kedes L (1986) Codon usage of histone gene families in higher eukaryotes reflects functional rather than phylogenetic relationships. J Mol Evol 23:224–241

Witt O, Albig W, Doenecke D (1996) Testis-specific expression of a novel human H3 histone gene. Exp Cell Res 229:301–306

Worcel A, Burgi E (1972) On the structure of the folded chromosome of *Escherichia coli*. J Mol Biol 71:127–147

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, Dyer KD, Rosenberg HF (2000) Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. Proc Natl Acad Sci USA 97:4701–4706

Zlatanova J (1997) Archaeal chromatin: virtual or real? Proc Natl Acad Sci USA 94:12251–12254