Evolution of High Mobility Group Nucleosome-Binding Proteins and Its Implications for Vertebrate Chromatin Specialization

Rodrigo González-Romero,[†,1] José M. Eirín-López,[†,2] and Juan Ausió[*,1]

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

[2]Chromatin Structure and Evolution (CHROMEVOL) Group, Department of Biological Sciences, Florida International University

[†]These two authors have contributed equally to this work.

***Corresponding author:** E-mail: jausio@uvic.ca.

Associate editor: Helen Piontkivska

Abstract

High mobility group (HMG)-N proteins are a family of small nonhistone proteins that bind to nucleosomes (N). Despite the amount of information available on their structure and function, there is an almost complete lack of information on the molecular evolutionary mechanisms leading to their exclusive differentiation. In the present work, we provide evidence suggesting that HMGN lineages constitute independent monophyletic groups derived from a common ancestor prior to the diversification of vertebrates. Based on observations of the functional diversification across vertebrate HMGN proteins and on the extensive silent nucleotide divergence, our results suggest that the long-term evolution of HMGNs occurs under strong purifying selection, resulting from the lineage-specific functional constraints of their different protein domains. Selection analyses on independent lineages suggest that their functional specialization was mediated by bursts of adaptive selection at specific evolutionary times, in a small subset of codons with functional relevance—most notably in HMGN1, and in the rapidly evolving HMGN5. This work provides useful information to our understanding of the specialization imparted on chromatin metabolism by HMGNs, especially on the evolutionary mechanisms underlying their functional differentiation in vertebrates.

Key words: **chromatin, high mobility group proteins, HMGN, nucleosome-binding domain, long-term evolution, purifying selection, episodic adaptive selection.**

Introduction

The high mobility group (HMG) proteins are the most abundant and ubiquitous nonhistone chromosomal proteins. They bind to DNA and to nucleosomes, eliciting structural changes on DNA metabolic activities such as transcription, replication, and DNA repair (Bustin and Reeves 1996; Bustin 1999). The HMG superfamily is composed of three families: HMGA, HMGB, and HMGN (Bustin 2001a), each containing a unique structural motif (Bianchi and Agresti 2005). The characteristic domains are: AT-hook for the HMGA family, the HMG Box for the HMGB family, and the nucleosome-binding domain (NBD) for the members of the HMGN family.

The first HMGN proteins (HMGN1 and HMGN2) were identified by E.W. Johns group under the names HMG-14 and HMG-17 (Johns 1982) (see [Bustin 2001b] for a change in HMG nomenclature). HMGNs are expressed only in vertebrates and interact with the 145-bp nucleosome core particle (NCP) (fig. 1*A*) without any DNA sequence specificity (Bustin 2001a). Binding of HMGNs to nucleosomes has downstream functional implications for transcription, replication, and repair (Vestner et al. 1998; Bustin 2001a; Birger et al. 2005; Belova et al. 2008; Kim et al. 2009; Postnikov and Bustin 2010; Zhu and Hansen 2010). Nucleosome binding is mediated by the NBD, a highly conserved 30 amino acid domain in vertebrates (Bustin 2001a) (fig. 1*A*). It contains an eight amino acid motif, RRSARLSA, responsible for the anchoring of HMGNs to the NCP (Ueda et al. 2008). Sumoylation of lysines 17 and 35 in HMGN2 within the NBD region has been recently shown to decrease the nucleosome-binding affinity of these proteins (Wu et al. 2014). In addition to the NBD, all HMGN proteins have a bipartite nuclear localization signal (Hock, Scheer, et al. 1998) and a negatively charged regulatory domain (RD) in their C-terminal region that mediates chromatin unfolding (Bustin 2001a) and plays an important role in the effects of HMGNs on histone posttranslational modifications (PTMs) (Ueda et al. 2006). The HMGN family consists of five closely related proteins that have been detected only in vertebrates: HMGN1, HMGN2, HMGN3, HMGN4, and HMGN5. Although HMGN1 to HMGN4 are approximately 100 amino acids in length, HMGN5 is much larger (300–400 amino acids) due to the presence of a long acidic C-terminal region (Rochman et al. 2010) that affects the cellular localization and architectural properties of the protein (Rochman et al. 2009).

HMGN1 and HMGN2 are the most abundant and characterized proteins of the HMGN family. In addition to nucleosome binding, they reduce the compactness of chromatin fiber, and enhance transcription of chromatin templates (Bustin 2001a). Both in vivo and in vitro studies suggest that they bind to NCPs, forming homodimeric complexes containing two molecules of either HMGN1 or HMGN2 (Postnikov et al. 1995, 1997). Using cross-linking and methyl-based NMR analysis, it has been demonstrated that these proteins unfold chromatin by targeting several of the

Article

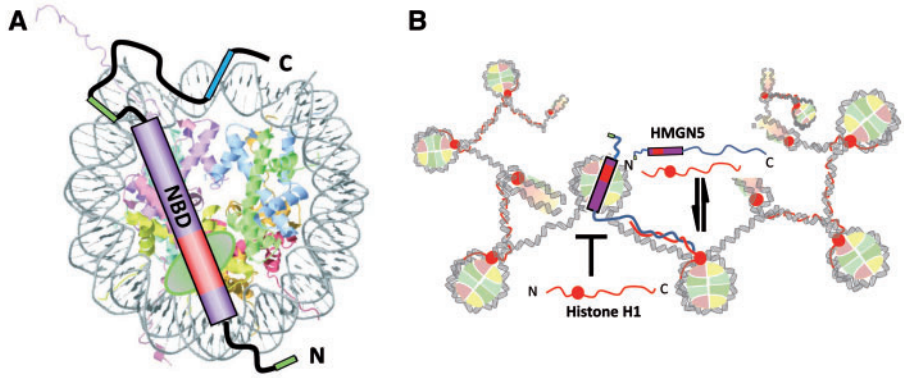

FIG. 1. Schematic representation of the interactions of HMGNs with chromatin. (*A*) Interaction of HMGN2 with the nucleosome (Kato et al. 2011; Kugler et al. 2012). The core histones are depicted in different light colors: H3: blue, H4: green, H2A: yellow, H2B pink. The green oval indicates the approximate location of the acidic patch (Luger et al. 1997). The colors for the HMGN2 molecule correspond to the different structural regions along its amino acid sequence (as indicated in fig. 2). Interaction of the NBD of HMGNs with the nucleosome positions their C-terminal domain near the nucleosome dyad. This results in an impairment of the proper binding of the winged histone domain (WHD) (Kasinsky et al. 2001) of linker histones to this region (Zhou et al. 2013). (*B*) Interaction of HMGN5 with chromatin results in a relaxed open conformation of the chromatin fiber, which prevents histone H1 binding. Such an unfolding stems from the binding competition between HMGN5 and histone H1 for the dyad region of the nucleosome, and/or from the juxtaposition of their respective negatively and positively charged C-terminal domains (Rochman et al. 2009, 2010). The red circle in the histone H1 molecule represents its highly characteristic WHD. The double arrow underscores the highly dynamic nature of the interactions of histone H1 and HMGNs with the chromatin template (Kugler et al. 2012).

main elements essential to chromatin compaction: the linker histone H1 and the N-terminal tails of histones H3 and H4 (Trieschmann et al. 1998; Catez et al. 2002; Kato et al. 2011). In addition, it has also been shown that both HMGN1 and HMGN2 can form multiple complexes with other nuclear proteins, which could alter their chromatin interaction and biological function (Lim et al. 2002; Kugler et al. 2012).

Although HMGN1 and HMGN2 seem to be highly expressed during embryogenesis, they are also ubiquitously expressed in several adult tissues (Furusawa and Cherukuri 2010). Immunofluorescence studies have pinpointed their numerous foci within the nucleus (Postnikov et al. 1997). Such organization appears to be highly dynamic and dependent on transcriptional activity (Hock, Wilde, et al. 1998). Binding of HMGN1 and HMGN2 proteins to nucleosomes is affected by PTMs such as phosphorylation and acetylation that can reduce or even abolish their nucleosome-binding ability (Bergel et al. 2000; Prymakowska-Bosak et al. 2001; Gerlitz et al. 2009; Pogna et al. 2010). Finally, several studies also suggest that both proteins can indirectly modulate the levels of some of the histone PTMs, hence affecting the histone-mediated epigenetic regulation of gene expression (Postnikov and Bustin 2010; Kugler et al. 2012).

HMGN3 (formerly Trip7 [thyroid hormone receptor interacting protein 7]) is the only family member to consist of two splice variants, HMGN3a and HMGN3b (West et al. 2001). The shorter HMGN3b variant lacks the C-terminal RD. During the course of this study, while searching in NCBI databases, two more splice variants were identified: we have named them HMGN3c and HMGN3d (see Results and Discussion section and supplementary table S1, Supplementary Material online). It remains to be determined whether any of these HMGN3 variants play a distinct role in vivo. Although HMGN1 and HMGN2 are ubiquitously expressed, and involved in general cellular differentiation, HMGN3 expression seems to be tissue-specific, and dependent upon development (Ueda et al. 2009). In mouse and human tissues, HMGN3 is highly expressed in the eye and in the brain, (West et al. 2001, 2004; Ito and Bustin 2002), where it might play a role in astrocyte function (Ito and Bustin 2002). Furthermore, HMGN3 is also abundant in adult pancreatic islet cells, where it modulates the transcriptional program of these cells, affecting insulin secretion (Ueda et al. 2009).

HMGN4 is the least-studied member of all HMGNs. Closely related to HMGN2, HMGN4 was identified in 2001 during a GenBank database-search of a new HMGN2-like transcript (Birger et al. 2001). In contrast to the rest of HMGNs, which are all encoded by genes containing six distinct exons, HMGN4 is encoded by an intronless gene (Birger et al. 2001). Also, although all the other HMGNs have been detected in all vertebrates that have been tested, the gene coding for HMGN4 appears to be restricted to primates (Kugler et al. 2012). The HMGN4 gene seems to have originated around 25 Ma from a fortuitous insertion of an HMGN2 retro-pseudogene next to an active promoter (Birger et al. 2001). This had been earlier recognized as a possibility (Srikantha et al. 1987), as HMGN is one of the largest known retro-pseudogene families, with human and mouse genomes containing more than 30 retro-pseudogenes for HMGN1 and HMGN2, dispersed over many chromosomes (Popescu et al. 1990; Johnson et al. 1992, 1993; Strichman-Almashanu et al. 2003). HMGN4 expression appears to be widespread among human tissues (with a higher expression in the thyroid gland, thymus, and lymph nodes), albeit with a transcript and protein abundance significantly lower than that of HMGN2 (Birger et al. 2001).

HMGN5 (previously known as NBP-45 [nucleosomal-binding protein 45], or NSBP1 [nucleosome-binding protein 1]) is the most recently described HMGN variant (Shirakawa et al. 2000; King and Francomano 2001; Rochman et al. 2009). It is a rapidly evolving protein that modulates the dynamic binding of linker histones (histone H1) to chromatin, reducing the compaction of the chromatin fiber, and affecting transcription. HMGN5 contains a long acidic C-terminal domain that differs among different vertebrate species (Malicet et al. 2011). The exon that encodes for this C-terminal region (exon VI) contains sequences highly similar to both HAL1 retro-transposable element and HERVH endogenous retrovirus; these similarities could be related to HMGN5's rapidly evolving nature (King and Francomano 2001; Malicet et al. 2011). The C-terminal region of HMGN5 is the main determinant of its chromatin interaction properties and its chromatin location (Rochman et al. 2009). For instance, mouse HMGN5 (with a 300 amino acid C terminus) (fig. 1*B*) is preferentially found in euchromatin, whereas human HMGN5 (with a 200 amino acid C terminus) exhibits a less restricted dual euchromatin and heterochromatin localization—similar to other HMGN variants (Malicet et al. 2011). Although its biological function is unknown, overexpression of HMGN5 has been observed in several human tumors, such as in prostate cancer (Jiang et al. 2010), squamous cell carcinoma (Green et al. 2006), renal cell carcinoma (Ji et al. 2012), breast cancer (Li et al. 2006), gliomas (Qu et al. 2011), and lung cancer (Chen et al. 2012); this suggests that HMGN5 plays a role in tumorigenesis. Knockdown of HMGN5 induces cell cycle arrest and apoptosis in these human tumor cell lines; it has thus been suggested that HMGN5 might be a potential molecular target for cancer therapy (Chen et al. 2012).

In the present work, we trace the phylogeny and evolutionary history of HMGNs, which led to their structural and functional specialization during the course of vertebrate evolution.

Results and Discussion

Vertebrate HMGN2 Distribution and Tissue Variation

As mentioned in the introduction, HMGN1/2 display a heterogeneous pattern of distribution and expression across vertebrates, an animal group within which they appear to have had their evolutionary emergence. Attempts to extract any similar proteins in invertebrate organisms were unsuccessful in both this and in previous studies (Bustin 2001a). Figure 2*A* shows the alignment of HMGN1 and HMGN2 in five species representative of each of the five classes within the subphylum vertebrata. The Logos representation shown underneath the amino acid sequences highlights the extent to which their different structural domains have been conserved.

To gain insight regarding the distribution of these HMGNs and their relative abundance, a 5% perchloric acid (PCA) extraction was performed on a liver sample from the same organisms used in the sequence alignments. This type of acid extraction not only extracts HMGNs, but it also extracts the linker histones of the histone H1 family (Goodwin et al. 1978). We took advantage of this dual extraction to produce an approximate normalization of the protein loadings for each extraction (fig. 2*B*) prior to performing the Western-blot analysis with an HMGN2 mouse antibody. Attempts to perform a similar analysis with HMGN1 and HMGN3 mouse antibodies proved to be completely unsuccessful. As shown in figure 2*B*, HMGN2 exhibits a variable distribution across the vertebrate species analyzed here, with a lower expression in chicken and an enhanced electrophoretic mobility in zebrafish—in agreement with the smaller size of the amino acid sequence in this organism (see fig. 2*A* and supplementary fig. S1, Supplementary Material online).

More striking is the variability that is observed across different tissues within the same organism, as exemplified by the Western-blot analysis carried out on mice and shown in figure 2*C*. The major occurrence of HMGN2 appears to be in the brain, followed by intestines and lungs, testes, kidneys, and liver. In partial agreement with these results, a previous study on the variation of HMGN2 in liver, kidney, and lung tissues of rats was not able to detect a significant variation within these tissues in this organism, but did consistently show a larger presence in lung tissue (Kuehl et al. 1984). Although the presence of HMGN2 in the nucleus has been related to the transcriptional activity of the cell (Hock, Wilde, et al. 1998), its relation to the tissue variability observed by us—and its potential significance—deserves further attention.

As mentioned above, attempts to extend our tissue and organism distribution analysis using mouse antibodies failed. Figure 3 provides an amino acid sequence analysis for this HMGN3 as well as for the primate-specific HMGN4 and HMGN5. In the absence of a Western-blot analysis tool, our bioinformatics search for HMGN3 occurrence only allowed us to detect the presence of this protein in mammals, in birds, and in *Xenopus*, but not in any other vertebrate species for which whole-genome information is available.

Phylogenetic Relationships among HMGN Family Members

The availability of complete information on many vertebrate organism genomes provides a unique opportunity to address a fundamental question as to how the different HMGNs correlate to each other throughout vertebrate evolution. To this end, protein and gene phylogenies were reconstructed from sequence data obtained after exhaustive molecular data mining (see supplementary table S1, Supplementary Material online). The resulting protein and gene phylogenies are shown in figure 4 and supplementary figure S2, Supplementary Material online, respectively. In both instances, the five major HMGN lineages, as well as the HMG-14A group, are well defined; each of them represents a distinctive monophyletic clade, as supported by the high confidence values observed. Given that the bootstrap (BS) method is known to be conservative, values higher than 80% were interpreted as high statistical support for internal nodes on both trees. The results were additionally supported by high Bayesian posterior probabilities in those branches leading to each HMGN lineage. Such a clustering pattern is

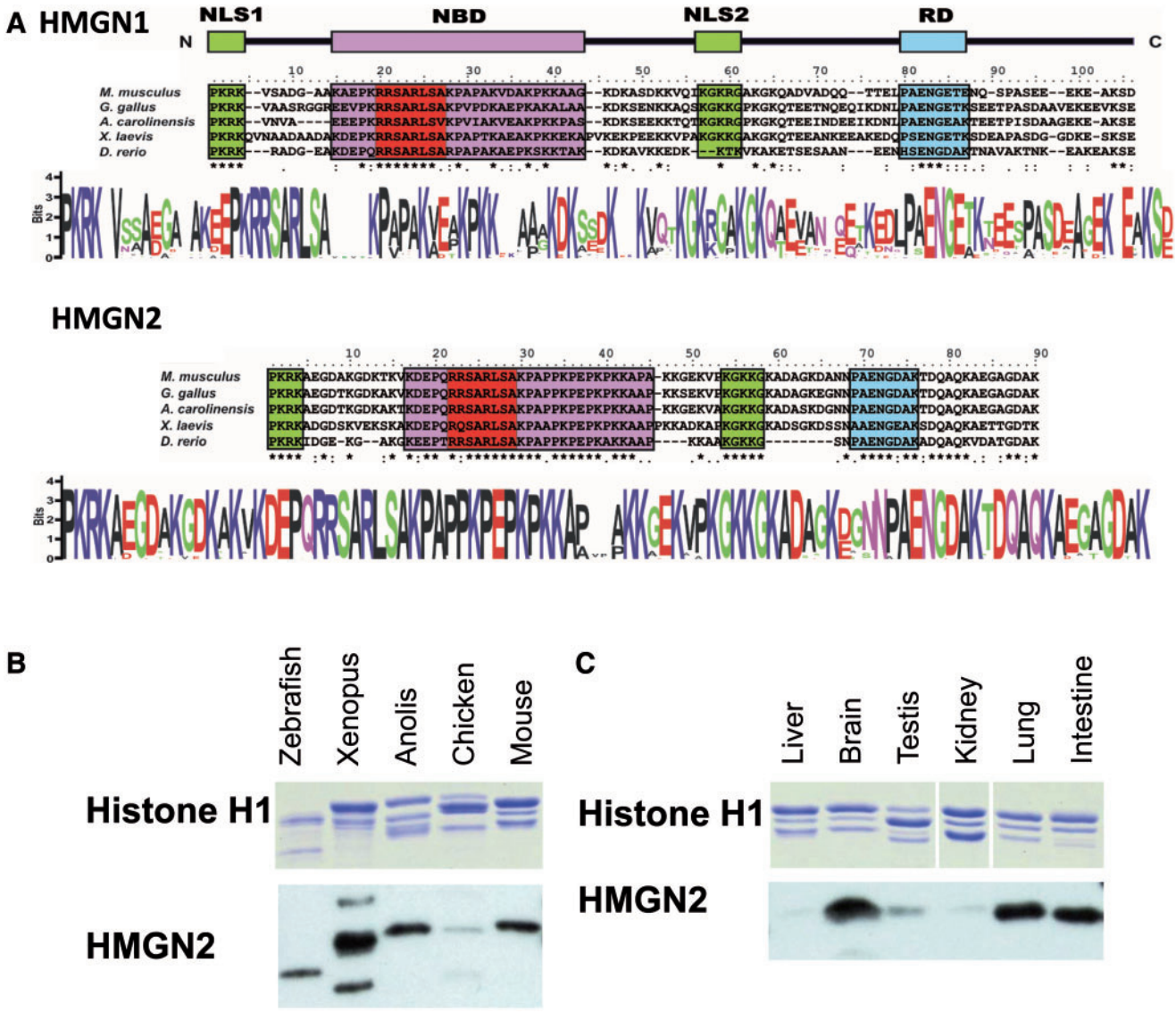

FIG. 2. HMGN1 and HMGN2. (*A*) Protein sequence alignment for a representative organism of each of the five vertebrate classes: Zebrafish, *Danio rerio* (fish); African clawed frog, *Xenopus laevis* (amphibian); Carolina anole, *Anolis carolinensis* (reptile); chicken, *Gallus gallus* (bird); and mouse, *Mus musculus* (mammal). The combined Logos representations, using alignments from supplementary figure S1, Supplementary Material online, are also shown. (*B*) Western-blot analysis of HMGN2 from liver tissue-PCA extracts from each one of the vertebrate representatives in (*A*). A coomassie blue-stained replica SDS–PAGE corresponding to the histone H1 fraction coextracted in this way is also shown. (*C*) Coomassie blue stained SDS–PAGE and Western-blot analysis of HMGN2 PCA extracted from different mouse tissues (liver, brain, testis, kidney, lung, and gut). In (*B*) and in (*C*), histones H1 were used for protein loading normalization purposes.

consistent with the presence of specific constraints acting upon different HMGN lineages, which leads to a functional diversification that is likely to have different downstream structural and functional implications for chromatin.

The reconstructed topologies support a retroviral origin of HMGN4 (from an HMGN2 retro-pseudogene [Birger et al. 2001]), as well as a close relationship between the bird/reptile HMG-14A group and HMGN3 (Browne and Dodgson 1993) (table 1). Unfortunately, actual sequence data does not allow us to indicate which HMGN is the closest one to a common ancestor. Although the confidence values obtained for internal nodes within the protein phylogeny allow us to discern beween monophyletic groups corresponding to each HMGN type, it is not possible to solve the deep relationships among HMGNs beyond each group, probably due to the accumulation of multiple substitutions at individual amino acid sites. However, the taxonomic distribution and the wide distribution of HMGNs across vertebrates suggest that HMGN1 and HMGN2 (the two founding members of the HMGN family) as well as HMGN3 arose earlier in evolution. In contrast, HMGN4 (present in catarrhini primates) and HMGN5 (present in mammals) appear to be the most recent lineages, originating 25 and 300 Ma, respectively (Birger et al. 2001; Malicet et al. 2011), with the latter corresponding to the most sequence-divergent as a result of its rapid evolution (Malicet et al. 2011) (table 1).

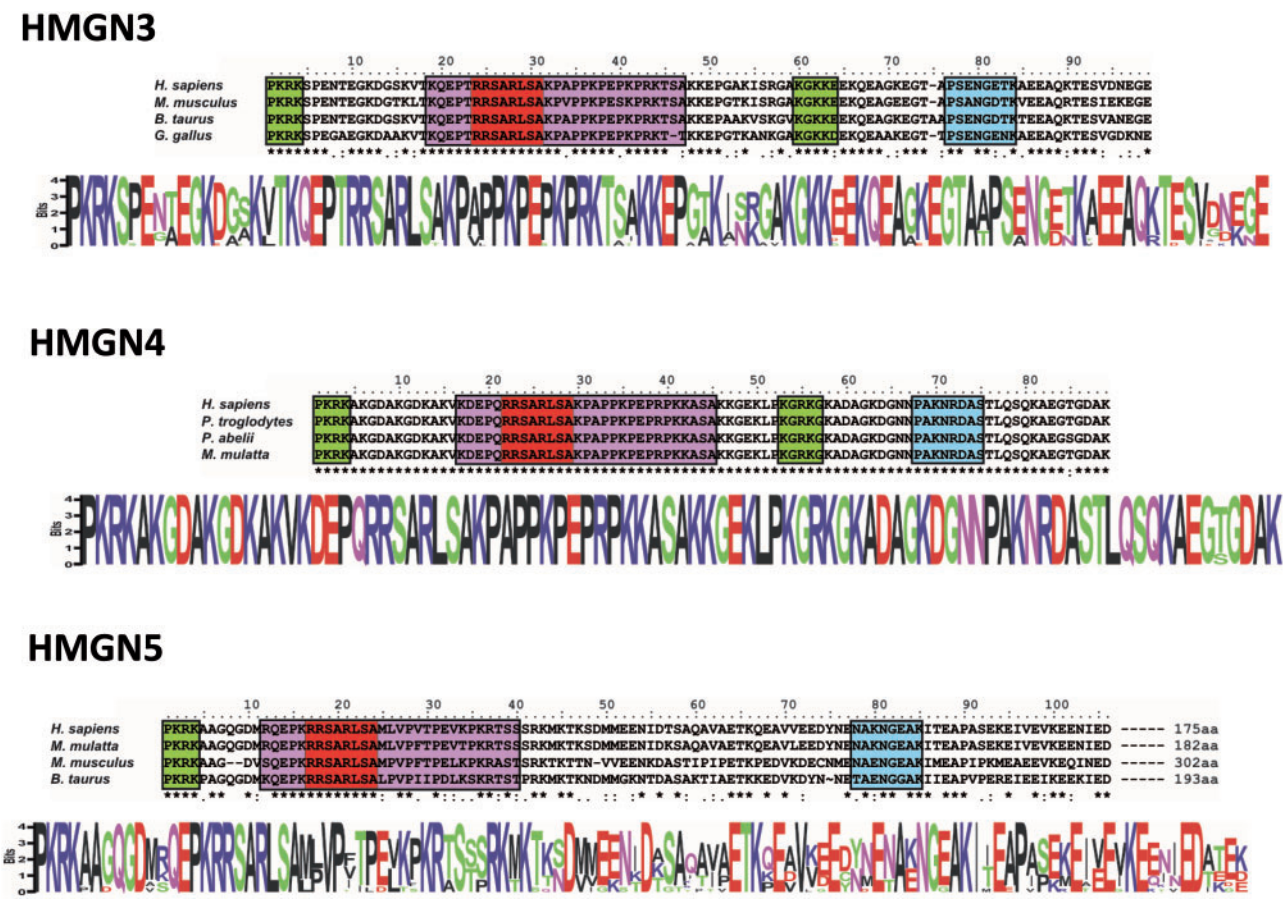

FIG. 3. HMGN3, HMGN4, and HMGN5. Protein sequence alignment for different representative organisms: chicken, *Gallus gallus*; cow, *Bos taurus*; mouse, *Mus musculus*; Rhesus macaque, *Macaca mulatta*; chimpanzee, *Pan troglodytes*; orangutan, *Pongo abelii*; human, *Homo sapiens*. The combined Logos representations, using alignments from supplementary figure S1, Supplementary Material online, are also shown.

Mechanisms of HMGN Evolution

The phylogenetic analysis shown in figure 4 depicts a highly specialized differentiation of HMGNs which is likely related to a functional specialization. Which, then, are the mechanisms that govern the long-term evolution of these different lineages? To address this question, we started by examining the protein variation within each of the different HMGN lineages. Such analysis revealed that HMGN5 is the most diverse ($p = 0.326 \pm 0.015$), followed by HMGN1, HMGN3, and HMGN2, with HMGN4 having the lowest levels of variation ($p = 0.004 \pm 0.004$) (table 2). The nature of the nucleotide variation underlying such diversity was predominantly synonymous and in all instances higher than the nonsynonymous variation. As expected, the lowest levels of silent variation were found in HMGN4 ($p_S = 0.055 \pm 0.020$), likely mirroring its recently retroposed origin (Birger et al. 2001; Strichman-Almashanu et al. 2003). Still, when it comes to complete proteins, codon-based *Z*-tests for selection consistently revealed significant differences between synonymous and nonsynonymous variation (table 2). Altogether, these results support the presence of a strong purifying selection operating on the different HMGN protein lineages—which is most likely responsible for preserving the structural features required for the specific interaction of each of these proteins with the nucleosome (Bustin 2001a).

Evidence for the role of purifying selection was further supported by the low levels of protein variation found at the N-terminal domain of HMGNs (table 2). This region probably represents the main target of selection, as it encompasses the most functionally relevant binding domain (NBD) for the interaction of these proteins with the nucleosome (Kato et al. 2011). Comparatively, the higher nonsilent variation found at the C-terminal region is probably due to a low selectivity for acidic amino acids in the regulatory domain (RD). This becomes especially evident in the long C-terminal region of HMGN5, which contains high levels of either aspartic or glutamic acid, organized in the repetitive motif EDGKE. The highly acidic nature of this domain represents the main determinant of the chromatin interaction properties of HMGN5 (fig. 1*B*) (Malicet et al. 2011), and it also plays a critical role in transcriptional regulation by modulating the occurrence of specific chromatin modifications (Ueda et al. 2006).

To test whether HMGN specialization hints at the involvement of additional lineage-specific functional constrains, we estimated the pace at which each HMGN lineage evolves. The analysis showed low-to-moderate rates of evolution in all instances—except in HMGN5, which appears to be evolving at a very fast rate (fig. 5). In this regard, HMGN5 represents a lineage with an outstandingly rapid rate of evolution, reminiscent of chromosomal reproductive proteins (Eirín-López

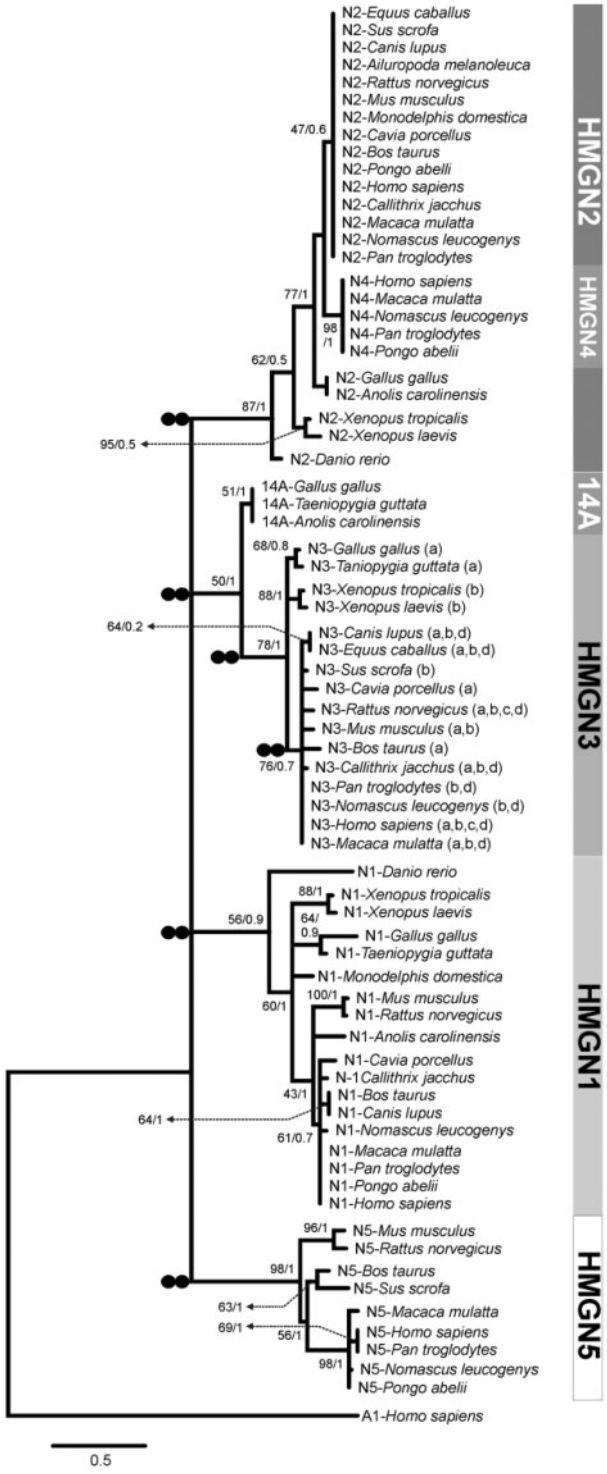

Fig. 4. Phylogenetic maximum likelihood (ML) relationships among vertebrate HMGN protein lineages. The numbers for interior branches represent nonparametric bootstrap (BS) probabilities based on 1,000 replications, followed by Bayesian posterior probabilities (only shown when BS $\geq$ 50% or posterior probability $\geq$ 0.5). Two black circles at internal nodes indicate subtrees at which the molecular clock hypothesis was rejected ($P < 0.001$) after testing for the presence of local molecular clocks.

et al. 2008; Ishibashi et al. 2010). Quite unexpectedly, such a high rate of evolution does not preclude the use of preferred codons in HMGN5 genes, as shown by the codon bias

Table 1. Evolutionary Divergence between HMGN Protein Lineages across Vertebrates[a].

	HMGN1	HMGN2	HMGN3	HMGN4	HMGN5
HMGN1	—				
HMGN2	24.9 ± 5.8	—			
HMGN3	28.2 ± 6.3	22.0 ± 6.3	—		
HMGN4	31.0 ± 6.5	8.8 ± 4.4	24.7 ± 6.9	—	
HMGN5	33.7 ± 7.2	38.1 ± 7.3	32.5 ± 6.8	38.3 ± 7.4	—

[a]Average amino acid substitutions per 100 sites (*p*-distance). Standard errors were calculated using the BS method with 1,000 replicates.

estimations (table 2). This would support the existence of specialized constraints in the evolution of HMGN5, which are different from those operating in other lineages.

Episodic Selection within HMGN Lineages

Despite all the consistent evidence for the importance of purifying selection in shaping the functional differentiation of HMGNs, the presence of heterogeneous evolutionary rates across lineages—together with the high level of divergence displayed by the recently differentiated HMGN5 lineage—raises the question as to whether or not there is any evidence for adaptive selective episodes driving the rapid differentiation of specific HMGN lineages. Should this be the case, it would be expected to have traces of these episodes detected across HMGN evolution. This notion is supported by our results, which show a significant departure from a global clock-like behavior during the evolution of HMGN proteins (lnL without clock = −5,998.0, lnL with clock = −2,1304.4, $P < 0.001$), resulting from heterogeneous rates of evolution at internal branches leading to the different HMGN lineages ($P < 0.001$, fig. 4).

Because HMGN5 lineage is only present in mammals, we decided to base our analysis on the evolution of HMGN genes in this group. Lineages HMGN2, HMGN4, and HMGN3 are closely related, within a single monophyletic group—with HMGN1 and the HMGN5 lineages constituting a sister clade (fig. 6A). As it was done for vertebrates, the global molecular clock hypothesis was also tested and rejected in the mammalian groups ($P < 0.001$), which exhibit a significant departure from a clock-like behavior found at the monophyletic origins of each HMGN clade (fig. 6A). Given the presence of heterogeneous rates of evolution, we investigated to what extent those resulted from specific selective episodes operating on particular HMGNs. The screening of the HMGN phylogeny revealed significant traces of episodic adaptive selection ($\omega > 1$) on at least five internal branches ($P \leq 0.05$) (fig. 6A). Although one of these branches is located at the root of the HMGN4 lineage, the four remaining branches are located in the subtree encompassing lineages HMGN1 and HMGN5, including the root of this clade ($P \leq 0.01$), as well as the internal branch leading to the HMGN1 lineage ($P \leq 0.01$), and the branches grouping murine ($P \leq 0.001$) and catarrhini ($P \leq 0.05$) HMGN5 genes together.

Table 2. Average Numbers of Amino Acid (p_{AA}), Nucleotide (p_{NT}), Synonymous (p_S), and Nonsynonymous (p_N) Nucleotide Differences per 100 Sites Site in HMGN Lineages, Discriminating among Complete Coding Regions, N-terminal and C-Terminal Domains[a].

HMGN Type	p_{AA} (SE)	p_{NT} (SE)	p_S (SE)	p_N (SE)	*R*	*Z*-test	ENC
HMGN1 complete	23.3 ± 2.4	22.1 ± 1.4	49.9 ± 2.4	13.0 ± 1.5	1.0	13.7**	50.1 ± 5.9
HMGN1 N-terminus	16.4 ± 3.6	20.9 ± 2.1	53.0 ± 2.5	9.9 ± 2.1	0.8	13.4**	51.7 ± 10.7
HMGN1 C-terminus	28.5 ± 3.4	22.9 ± 1.8	47.1 ± 3.4	15.3 ± 2.0	1.2	8.0**	44.6 ± 8.7
HMGN2 complete	6.8 ± 1.3	10.7 ± 1.0	30.4 ± 2.5	3.5 ± 0.7	1.7	10.1**	45.3 ± 7.6
HMGN2 N-terminus	6.3 ± 1.8	11.4 ± 1.4	32.5 ± 3.4	3.4 ± 0.9	1.6	8.2**	48.8 ± 6.2
HMGN2 C-terminus	7.5 ± 1.9	9.8 ± 1.5	27.3 ± 3.8	3.6 ± 1.0	1.7	5.8**	49.7 ± 0.0
HMGN3 complete	8.7 ± 1.8	9.6 ± 1.0	22.9 ± 2.3	4.3 ± 0.9	1.5	7.7**	43.1 ± 5.7
HMGN3 N-terminus	7.7 ± 2.0	9.5 ± 1.1	24.5 ± 2.6	3.8 ± 1.0	1.4	7.5**	49.6 ± 6.6
HMGN3 C-terminus	10.4 ± 3.0	9.7 ± 1.7	20.0 ± 4.1	5.1 ± 1.5	1.7	3.1**	43.1 ± 7.8
HMGN4 complete	0.4 ± 0.4	1.5 ± 0.5	5.5 ± 2.0	0.2 ± 0.2	0.7	2.6*	47.0 ± 1.3
HMGN4 N-terminus	0.0 ± 0.0	1.7 ± 0.8	6.6 ± 3.0	0.0 ± 0.0	1.0	2.2*	42.5 ± 0.0
HMGN4 C-terminus	2.9 ± 0.9	1.2 ± 0.7	3.9 ± 2.6	0.4 ± 0.4	0.6	1.2	39.8 ± 0.0
HMGN5 complete	32.6 ± 1.5	19.4 ± 0.8	23.1 ± 01.6	18.3 ± 1.0	1.4	2.7*	39.9 ± 2.3
HMGN5 N-terminus	16.2 ± 3.5	9.8 ± 1.7	13.9 ± 03.9	8.3 ± 1.9	1.7	1.2	34.9 ± 6.4
HMGN5 C-terminus	35.6 ± 1.7	21.2 ± 0.8	25.2 ± 01.7	20.0 ± 1.1	1.4	2.7*	38.6 ± 3.4

Note.—SE, standard error; ENC, Effective Number of Codons (codon bias) ranging between 61 (no bias) and 20 (maximum bias). HMGN1: N-terminus nucleotide positions 1–153, C-terminus positions 154–342; HMGN2 N-terminus positions 1–147, C-terminus positions 148–279; HMGN3 N-terminus positions 1–147, C-terminus positions 148–396; HMGN4 N-terminus positions 1–141, C-terminus positions 142–273; HMGN5 N-terminus positions 1–126, C-terminus positions 127–1314 (see Materials and Methods section for a detailed explanation).

[a]The average transition/transversion ratio used in the estimation of p_S and p_N is denoted as *R*. SEs calculated by the bootstrap method with 1,000 replicates.

*$P < 0.05$ and **$P < 0.001$ level in *Z*-test comparisons ($p_S > p_N$).

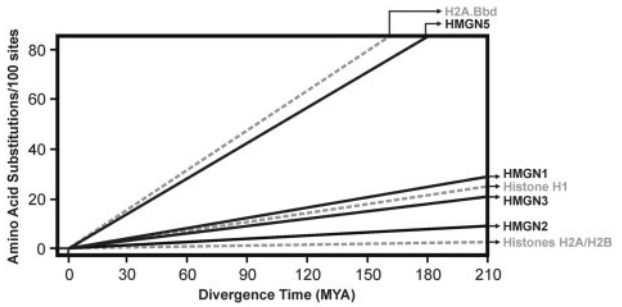

Fig. 5. Estimated rates of evolution for HMGN proteins. Evolutionary rates for the fast-evolving chromosomal proteins histone H2A.Bbd, as well as histone H1 and histones H2A/H2B (dashed lines) are included as reference. HMGN4 is not shown, due to its very slow rate of evolution.

Additional insight was gained by combining maximum likelihood (ML) and Bayesian selection analyses, which allowed us to disclose the individual sites subject to diversifying selection (Kosakovsky Pond and Frost 2005). As a result, 12 positively selected and 134 negatively selected sites were identified, based on the consensus of single-likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), random effects likelihood (REL), and fast unconstrained Bayesian approximation (FUBAR) methods (table 3). Among them, seven positively selected codons were consistently identified as subject to episodic positive selection based on the mixed effects model of evolution (MEME) method ($P < 0.1$), including three positions common to all HMGN types (49, 53, and 97) and four positions exclusively from the long C-terminal region of the HMGN5 lineage (table 3 and fig. 6*B*). The phylogenetic analysis of the mutations at these positions suggests that changes at codons 53 and 97 were most likely involved in the differentiation of the HMGN1 lineage, with changes in positions 49 and 53 linked to HMGN5. Interestingly, the presence of episodic selection at position 49 could constitute a major driver of HMGN5 specialization, given the location of this codon within the highly conserved and functionally relevant NBD region. Nonetheless, the differentiation of this latter lineage also required additional substitutions at positions 135, 363, 431, and 433 (fig. 6*B*).

Conclusions

HMGNs are characterized by their heterogeneous pattern of distribution and expression across vertebrates and have critical functions in chromatin metabolism. Yet, the evolutionary mechanisms responsible for such diversification and for the functional differentiation across their family members have eluded study. In the present work, we provide the first comprehensive analysis of the evolution of HMGNs, supplying evidence for three previously unknown major findings: 1) phylogenetic relationships among HMGN lineages show that all of them are independent monophyletic groups arising from a common ancestor that preceded the diversification of vertebrates; 2) long-term evolution of HMGNs is predominantly driven by purifying selection resulting from lineage-specific functional constraints of their different protein domains; 3) functional specialization of the different HMGN lineages occurred by bursts of adaptive selection at specific evolutionary times and protein positions, most notably in HMGN1 and in the rapidly evolving HMGN5. Altogether, our results suggest that HMGN evolution involves a heterogeneous process largely shaped by strong purifying selection, with occasional episodes of diversifying selection geared toward the functional specialization of the different lineages.

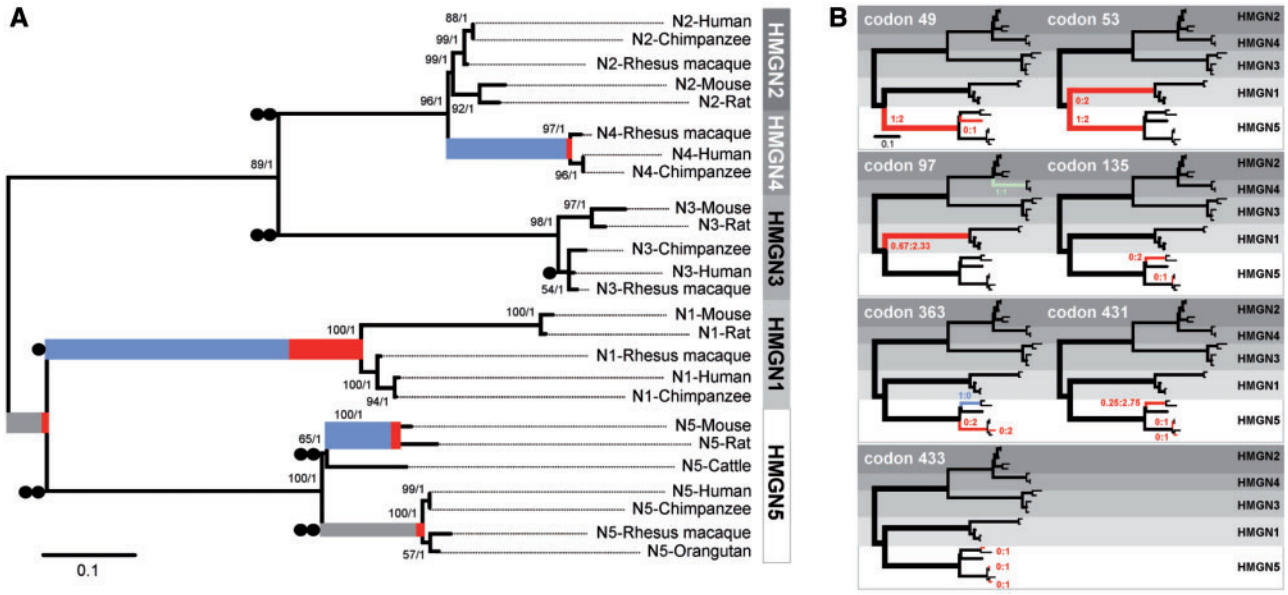

Fig. 6. Selection episodes involved in the evolution of mammalian HMGN lineages. (*A*) ML gene tree depicting episodes of diversifying selection during HMGN differentiation in mammals. Numbers for interior branches are indicated as in figure 4. Deviations from the molecular clock at internal subtrees are indicated by one ($P < 0.01$) or two ($P < 0.001$) black circles at the corresponding internal braches. The strength of selection at significant branches is represented in red ($\omega > 5$), gray ($\omega = 1$), and blue ($\omega = 0$), with the proportion of sites within each class represented by the color width. Thicker branches have been classified as undergoing episodic diversifying selection at corrected $P \leq 0.001$ (thickest branches), $P \leq 0.01$ (medium thickness), and $P \leq 0.05$ (thin branches). (*B*) Phylogenetic location of mutations involved in diversifying selection episodes during the evolution of HMGN genes. Branches in red account for higher numbers of nonsynonymous mutations, whereas branches in blue indicate higher numbers of synonymous mutations, and branches in green represent cases with equal numbers of nonsynonymous and synonymous mutations. Codon 49 is located within the highly conserved NDB region.

Materials and Methods

Extraction and Analysis of Distribution of HMGN Proteins

HMGN proteins were isolated from liver tissue of different vertebrate representatives, including: Fish (zebrafish *Danio rerio*), amphibian (African clawed frog *Xenopus laevis*), reptile (Carolina anole *Anolis carolinensis*), bird (chicken *Gallus gallus*), and mammalian (mouse *Mus musculus*) representatives. In addition, HMGNs were also extracted from several mouse tissues, including: brain, testis, kidney, lung, and intestine—as described elsewhere (Lim et al. 2004). Briefly, the tissues were processed with a dounce homogenizer in 0.15 M NaCl, 10 mM Tris-HCl (pH 7.5), and a 0.5% Triton X-100 buffer containing Roche Complete Protease cocktail inhibitor (Roche Molecular Biochemicals, Laval, QC) at a ratio of 1:100 v/v. After homogenization and incubation on ice for 5 min, the samples were centrifuged at 12,000 × g for 10 min at 4 °C. The resulting pellets were resuspended in 5% PCA, homogenized as above, and centrifuged in the same way. 1 N HCl was added to the PCA supernatant extracts to bring the solution to 0.2 N HCl. Then, the PCA supernatant extracts were precipitated with six volumes of acetone at −20°C overnight and centrifuged at 12,000 × g for 10 min at 4 °C. The acetone pellets were dried using a speed-vac concentrator, and stored at −80°C until further use in polyacrylamide gel electrophoresis (PAGE) and Western-blot analyses.

Gel Electrophoresis and Western Blotting

Sodium dodecyl sulfate (SDS)–PAGE (15% acrylamide, 0.4% bis-acrylamide) was carried out using the approach described by Laemmli (Laemmli and Johnson 1973). Western-blot analyses were performed using a mouse anti-HMGN2 antibody (a generous gift from Michael Bustin). Gels were electro-transferred to a polyvinylidene difluoride membrane (Bio-Rad, Hercules, CA) and processed as described elsewhere (Finn et al. 2008). HMGN2 antibody was used at a 1:2,000 dilution. Membranes were incubated with secondary goat antirabbit antibody (GE Healthcare, Baie d'Urfe, QC) at a 1:5,000 dilution. Secondary antibody was detected with enhanced chemiluminescence (GE Healthcare) and exposure to X-ray films.

Molecular Data Mining

Extensive data mining experiments were performed in the GenBank database (www.ncbi.nlm.nih.gov/genbank) in order to collect all the HMGN sequences available as of January 2014. Altogether, 88 nt coding sequences belonging to 21 different vertebrate species were used in the present work, including 18 HMGN1, 20 HMGN2, 33 HMGN3, 5 HMGN4, 9 HMGN5, 3 HMG-14A, and 1 outgroup sequence (HMGA1 from human, see supplementary table S1, Supplementary Material online). Sequences were revised for errors in accession numbers and nomenclature, and given that the HMGN family is one of the largest known retro-pseudogene families (Strichman-Almashanu et al. 2003), only functional HMGN coding sequences were selected. Multiple

Table 3. Codon Positions Potentially Subject to Selection during HMGN Evolution in Mammals[a].

Codon	SLAC (*P* value)	FEL (*P* value)	REL (Bayes factor)	FUBAR (posterior probability)	MEME (*P* value)
49	0.687	0.783	0.002	0.446	0.009*
53	0.491	0.552	0.006	0.563	0.038*
97	0.722	0.786	0.003	0.316	0.006*
128	0.000	0.086*	6.582	0.767	0.103
135	0.000	0.096*	7.228	0.750	0.076*
278	0.000	0.195	77.197*	0.950*	0.225
185	0.000	0.111	109.453*	0.920*	0.146
196	0.000	0.451	72.258*	0.821	0.448
363	0.000	0.082*	52.196*	0.942*	0.031*
376	0.000	0.460	53.381*	0.801	0.423
431	0.000	0.233	57.856*	0.853	0.025*
433	0.000	0.140	8.671	0.802	0.037*

[a]Positions subject to selection (*) as identified by the codon-based ML methods used to estimate ω at different positions.

sequence alignments were conducted on the basis of the translated amino acid sequences and edited for potential errors using the BIOEDIT (Hall 1999) and ClustalW (Thompson et al. 1994) programs. The alignment of the complete set of sequences consisted of 1,395 nt positions corresponding to 465 amino acid sites (supplementary fig. S1, Supplementary Material online). The boundaries of N-terminal (including the NBD) and acidic C-terminal regions of HMGN proteins (containing the RD) were established on the basis of the information available in literature as follows: HMGN1 N-terminus nucleotide positions 1–153, C-terminus positions 154–342 (Ding et al. 1997); HMGN2 N-terminus positions 1–147, C-terminus positions 148–279 (Crippa et al. 1992); HMGN3 N-terminus positions 1–147, C-terminus positions 148–396 (West et al. 2001); HMGN4 N-terminus positions 1–141, C-terminus positions 142–273 (Birger et al. 2001); HMGN5 N-terminus positions 1–126, C-terminus positions 127–1314 (King and Francomano 2001).

Phylogenetic Analysis of HMGNs

Molecular evolutionary analyses were performed using the computer program MEGA version 6 (Tamura et al. 2013), except where noted. Due to their smaller variance (Nei and Kumar 2000), nucleotide and protein sequence divergence was estimated using uncorrected differences (*p*-distances, partial deletion 95%). 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, and estimating standard errors by using the bootstrap (BS) method (1,000 replicates). HMGN phylogenies were reconstructed following a maximum likelihood (ML) approach, with the substitution models that best fit the analyzed sequences being JTT (Jones et al. 1992) and TN93 (Tamura and Nei 1993) including gamma-distributed variation across sites for protein and nucleotide sequences, respectively. Additional HMGN phylogenies were inferred in mammals (the only group in which all five HMGN lineages are represented), including: Human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), orangutan (*Pongo abelii*), rhesus macaque (*Macaca mulatta*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), and cow (*Bos taurus*). The reliability of the reconstructed topologies was contrasted in each case by nonparametric BS (1,000 replicates), and further examined by bayesian analysis using the program BEAST version 1.7 (Drummond et al. 2012), producing posterior probabilities. Three independent Markov chain Monte Carlo runs of 10,000,000 generations each were performed to generate posterior probabilities, sampling tree topologies every 1,000 generations to ensure the independence of successive trees, and discarding the first 1,000 trees of each run as burn-in. Trees were rooted with the human HMGA1a, a HMG protein functionally unrelated to HMGNs (Friedmann et al. 1993).

Molecular Evolution and Selection Analyses

The footprint of selection on HMGN genes was studied using two major approaches. First, descriptive analyses of nucleotide variation and the mode of evolution displayed by HMGNs were carried out. Accordingly, the numbers of synonymous (p_S) and nonsynonymous (p_N) nucleotide differences per site were compared using codon-based *Z*-tests for selection, setting 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). Additionally, the amount of codon usage bias and the presence of global and local molecular clocks were investigated using the programs DnaSP version 5 (Librado and Rozas 2009) and HyPhy (Pond et al. 2005), respectively. Finally, the rates of evolution of different HMGN lineages were estimated by correlating pairwise protein divergences between pairs of taxa with their corresponding divergence as defined by the TimeTree database (Hedges et al. 2006) (see supplementary table S2, Supplementary Material online). Regression analyses were implemented using the program STATGRAPHICS Plus version 5.1 (Warrenton, VA).

Second, the presence of lineages displaying evidence of diversifying (adaptive) selection episodes ($\omega > 1$) was examined across HMGN evolution by using the branch-site REL model (Pond and Frost 2005). To this end, a total of 444 codon positions were examined using an ML phylogeny that was reconstructed using HMGN nucleotide coding regions as a reference (in this instance, the best-fit model of evolution was defined as TN93 + G); no prior assumptions about which lineages have been subject to diversifying selection were made. The proportion of sites inferred to be evolving under diversifying selection at each branch were estimated using likelihood ratio tests, resulting in a *P* value for episodic selection. The strength of selection was partitioned for descriptive purposes into three categories ($\omega > 5$, $\omega = 1$, $\omega = 0$), using three different significance levels ($P < 0.001$, $P < 0.01$, and $P < 0.05$) to assess the obtained results. Additionally, the presence of selection at individual sites was assessed by using different codon-based ML methods including SLAC, FEL, REL, FUBAR, and MEME, with this latter one modeling variable ω (dN/dS) across lineages at

an individual site (Murrell et al. 2012). A total of seven codons subject to significant episodes of diversifying selection ($P < 0.05$) were detected using MEME, and analyzed in the context of the HMGN phylogeny, providing information on internal branches accumulating higher numbers of nonsynonymous mutations. All analyses in this section were carried out using the HyPhy program (Pond et al. 2005) and the Datamonkey webserver (Poon et al. 2009; Delport et al. 2010).

Supplementary Material

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

Acknowledgments

This work was supported by a Canadian Institutes of Health Research (CIHR) grant (MOP-97878) to J.A. R.G.-R. is the recipient of a postdoctoral fellowship from the Spanish Ministry of Education. J.M.E.-L. has been supported by a start-up grant from the College of Arts and Sciences at Florida International University (CAS-FIU).

References

Belova GI, Postnikov YV, Furusawa T, Birger Y, Bustin M. 2008. Chromosomal protein HMGN1 enhances the heat shock-induced remodeling of Hsp70 chromatin. *J Biol Chem.* 283:8080–8088.

Bergel M, Herrera JE, Thatcher BJ, Prymakowska-Bosak M, Vassilev A, Nakatani Y, Martin B, Bustin M. 2000. Acetylation of novel sites in the nucleosomal binding domain of chromosomal protein HMG-14 by p300 alters its interaction with nucleosomes. *J Biol Chem.* 275: 11514–11520.

Bianchi ME, Agresti A. 2005. HMG proteins: dynamic players in gene regulation and differentiation. *Curr Opin Genet Dev.* 15:496–506.

Birger Y, Catez F, Furusawa T, Lim JH, Prymakowska-Bosak M, West KL, Postnikov YV, Haines DC, Bustin M. 2005. Increased tumorigenicity and sensitivity to ionizing radiation upon loss of chromosomal protein HMGN1. *Cancer Res.* 65:6711–6718.

Birger Y, Ito Y, West KL, Landsman D, Bustin M. 2001. HMGN4, a newly discovered nucleosome-binding protein encoded by an intronless gene. *DNA Cell Biol.* 20:257–264.

Browne DL, Dodgson JB. 1993. The gene encoding chicken chromosomal protein HMG-14a is transcribed into multiple mRNAs. *Gene* 124:199–206.

Bustin M. 1999. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol Cell Biol.* 19:5237–5246.

Bustin M. 2001a. Chromatin unfolding and activation by HMGN(*) chromosomal proteins. *Trends Biochem Sci.* 26:431–437.

Bustin M. 2001b. Revised nomenclature for high mobility group (HMG) chromosomal proteins. *Trends Biochem Sci.* 26:152–153.

Bustin M, Reeves R. 1996. High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function. *Prog Nucleic Acid Res Mol Biol.* 54:35–100.

Catez F, Brown DT, Misteli T, Bustin M. 2002. Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO Rep.* 3:760–766.

Chen P, Wang XL, Ma ZS, Xu Z, Jia B, Ren J, Hu YX, Zhang QH, Ma TG, Yan BD, et al. 2012. Knockdown of HMGN5 expression by RNA interference induces cell cycle arrest in human lung cancer cells. *Asian Pac J Cancer Prev.* 13:3223–3228.

Crippa MP, Alfonso PJ, Bustin M. 1992. Nucleosome core binding region of chromosomal protein HMG-17 acts as an independent functional domain. *J Mol Biol.* 228:442–449.

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.

Ding HF, Bustin M, Hansen U. 1997. Alleviation of histone H1-mediated transcriptional repression and chromatin compaction by the acidic activation region in chromosomal protein HMG-14. *Mol Cell Biol.* 17: 5843–5855.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29: 1969–1973.

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.

Finn RM, Browne K, Hodgson KC, Ausio J. 2008. sNASP, a histone H1-specific eukaryotic chaperone dimer that facilitates chromatin assembly. *Biophys J.* 95:1314–1325.

Friedmann M, Holth LT, Zoghbi HY, Reeves R. 1993. Organization, inducible-expression and chromosome localization of the human HMG-I(Y) nonhistone protein gene. *Nucleic Acids Res.* 21: 4259–4267.

Furusawa T, Cherukuri S. 2010. Developmental function of HMGN proteins. *Biochim Biophys Acta.* 1799:69–73.

Gerlitz G, Hock R, Ueda T, Bustin M. 2009. The dynamics of HMG protein-chromatin interactions in living cells. *Biochem Cell Biol.* 87: 127–137.

Goodwin GH, Walker JM, Johns EW. 1978. Studies on the degradation of high mobility group non-histone chromosomal proteins. *Biochim Biophys Acta.* 519:233–242.

Green J, Ikram M, Vyas J, Patel N, Proby CM, Ghali L, Leigh IM, O'Toole EA, Storey A. 2006. Overexpression of the Axl tyrosine kinase receptor in cutaneous SCC-derived cell lines and tumours. *Br J Cancer.* 94: 1446–1451.

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, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22:2971–2972.

Hock R, Scheer U, Bustin M. 1998. Chromosomal proteins HMG-14 and HMG-17 are released from mitotic chromosomes and imported into the nucleus by active transport. *J Cell Biol.* 143:1427–1436.

Hock R, Wilde F, Scheer U, Bustin M. 1998. Dynamic relocation of chromosomal protein HMG-17 in the nucleus is dependent on transcriptional activity. *Embo J.* 17:6992–7001.

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.

Ito Y, Bustin M. 2002. Immunohistochemical localization of the nucleosome-binding protein HMGN3 in mouse brain. *J Histochem Cytochem.* 50:1273–1275.

Ji SQ, Yao L, Zhang XY, Li XS, Zhou LQ. 2012. Knockdown of the nucleosome binding protein 1 inhibits the growth and invasion of clear cell renal cell carcinoma cells in vitro and in vivo. *J Exp Clin Cancer Res.* 31:22.

Jiang N, Zhou LQ, Zhang XY. 2010. Downregulation of the nucleosome-binding protein 1 (NSBP1) gene can inhibit the in vitro and in vivo proliferation of prostate cancer cells. *Asian J Androl.* 12:709–717.

Johns EW 1982. The HMG chromosomal proteins. New York: Academic Press.

Johnson KR, Cook SA, Bustin M, Davisson MT. 1992. Genetic mapping of the murine gene and 14 related sequences encoding chromosomal protein HMG-14. *Mamm Genome.* 3:625–632.

Johnson KR, Cook SA, Ward-Bailey P, Bustin M, Davisson MT. 1993. Identification and genetic mapping of the murine gene and 20 related sequences encoding chromosomal protein HMG-17. *Mamm Genome.* 4:83–89.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 8: 275–282.

Kasinsky HE, Lewis JD, Dacks JB, Ausió J. 2001. Origin of H1 linker histones. *FASEB J.* 15:34–42.
Kato H, van Ingen H, Zhou BR, Feng H, Bustin M, Kay LE, Bai Y. 2011. Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. *Proc Natl Acad Sci U S A.* 108:12283–12288.
Kim YC, Gerlitz G, Furusawa T, Catez F, Nussenzweig A, Oh KS, Kraemer KH, Shiloh Y, Bustin M. 2009. Activation of ATM depends on chromatin interactions occurring before induction of DNA damage. *Nat Cell Biol.* 11:92–96.
King LM, Francomano CA. 2001. Characterization of a human gene encoding nucleosomal binding protein NSBP1. *Genomics* 71: 163–173.
Kosakovsky Pond SL, Frost SD. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol.* 22:1208–1222.
Kuehl L, Salmond B, Tran L. 1984. Concentrations of high-mobility-group proteins in the nucleus and cytoplasm of several rat tissues. *J Cell Biol.* 99:648–654.
Kugler JE, Deng T, Bustin M. 2012. The HMGN family of chromatin-binding proteins: dynamic modulators of epigenetic processes. *Biochim Biophys Acta.* 1819:652–656.
Laemmli UK, Johnson RA. 1973. Maturation of the head of bacteriophage T4. II. Head-related, aberrant tau-particles. *J Mol Biol.* 80: 601–611.
Li DQ, Hou YF, Wu J, Chen Y, Lu JS, Di GH, Ou ZL, Shen ZZ, Ding J, Shao ZM. 2006. Gene expression profile analysis of an isogenic tumour metastasis model reveals a functional role for oncogene AF1Q in breast cancer metastasis. *Eur J Cancer.* 42:3274–3286.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
Lim JH, Bustin M, Ogryzko VV, Postnikov YV. 2002. Metastable macromolecular complexes containing high mobility group nucleosome-binding chromosomal proteins in HeLa nuclei. *J Biol Chem.* 277: 20774–20782.
Lim JH, Catez F, Birger Y, Postnikov YV, Bustin M. 2004. Preparation and functional analysis of HMGN proteins. *Methods Enzymol.* 375: 323–342.
Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389:251–260.
Malicet C, Rochman M, Postnikov Y, Bustin M. 2011. Distinct properties of human HMGN5 reveal a rapidly evolving but functionally conserved nucleosome binding protein. *Mol Cell Biol.* 31:2742–2755.
Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764.
Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press.
Pogna EA, Clayton AL, Mahadevan LC. 2010. Signalling to chromatin through post-translational modifications of HMGN. *Biochim Biophys Acta.* 1799:93–100.
Pond SL, Frost SD. 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol.* 22: 478–485.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
Poon AF, Frost SD, Pond SL. 2009. Detecting signatures of selection from DNA sequences using Datamonkey. *Methods Mol Biol.* 537:163–183.
Popescu N, Landsman D, Bustin M. 1990. Mapping the human gene coding for chromosomal protein HMG-17. *Hum Genet.* 85:376–378.
Postnikov Y, Bustin M. 2010. Regulation of chromatin structure and function by HMGN proteins. *Biochim Biophys Acta.* 1799:62–68.
Postnikov YV, Herrera JE, Hock R, Scheer U, Bustin M. 1997. Clusters of nucleosomes containing chromosomal protein HMG-17 in chromatin. *J Mol Biol.* 274:454–465.
Postnikov YV, Trieschmann L, Rickers A, Bustin M. 1995. Homodimers of chromosomal proteins HMG-14 and HMG-17 in nucleosome cores. *J Mol Biol.* 252:423–432.
Prymakowska-Bosak M, Misteli T, Herrera JE, Shirakawa H, Birger Y, Garfield S, Bustin M. 2001. Mitotic phosphorylation prevents the binding of HMGN proteins to chromatin. *Mol Cell Biol.* 21: 5169–5178.
Qu J, Yan R, Chen J, Xu T, Zhou J, Wang M, Chen C, Yan Y, Lu Y. 2011. HMGN5:a potential oncogene in gliomas. *J Neurooncol.* 104: 729–736.
Rochman M, Malicet C, Bustin M. 2010. HMGN5/NSBP1: a new member of the HMGN protein family that affects chromatin structure and function. *Biochim Biophys Acta.* 1799:86–92.
Rochman M, Postnikov Y, Correll S, Malicet C, Wincovitch S, Karpova TS, McNally JG, Wu X, Bubunenko NA, Grigoryev S, et al. 2009. The interaction of NSBP1/HMGN5 with nucleosomes in euchromatin counteracts linker histone-mediated chromatin compaction and modulates transcription. *Mol Cell.* 35:642–656.
Shirakawa H, Landsman D, Postnikov YV, Bustin M. 2000. NBP-45, a novel nucleosomal binding protein with a tissue-specific and developmentally regulated expression. *J Biol Chem.* 275:6368–6374.
Srikantha T, Landsman D, Bustin M. 1987. Retropseudogenes for human chromosomal protein HMG-17. *J Mol Biol.* 197:405–413.
Strichman-Almashanu LZ, Bustin M, Landsman D. 2003. Retroposed copies of the HMG genes: a window to genome dynamics. *Genome Res.* 13:800–812.
Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 10:512–526.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.
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. *Nucl. Acids Res.* 22:4673–4680.
Trieschmann L, Martin B, Bustin M. 1998. The chromatin unfolding domain of chromosomal protein HMG-14 targets the N-terminal tail of histone H3 in nucleosomes. *Proc Natl Acad Sci U S A.* 95: 5468–5473.
Ueda T, Catez F, Gerlitz G, Bustin M. 2008. Delineation of the protein module that anchors HMGN proteins to nucleosomes in the chromatin of living cells. *Mol Cell Biol.* 28:2872–2883.
Ueda T, Furusawa T, Kurahashi T, Tessarollo L, Bustin M. 2009. The nucleosome binding protein HMGN3 modulates the transcription profile of pancreatic beta cells and affects insulin secretion. *Mol Cell Biol.* 29:5264–5276.
Ueda T, Postnikov YV, Bustin M. 2006. Distinct domains in high mobility group N variants modulate specific chromatin modifications. *J Biol Chem.* 281:10182–10187.
Vestner B, Bustin M, Gruss C. 1998. Stimulation of replication efficiency of a chromatin template by chromosomal protein HMG-17. *J Biol Chem.* 273:9409–9414.
West KL, Castellini MA, Duncan MK, Bustin M. 2004. Chromosomal proteins HMGN3a and HMGN3b regulate the expression of glycine transporter 1. *Mol Cell Biol.* 24:3747–3756.
West KL, Ito Y, Birger Y, Postnikov Y, Shirakawa H, Bustin M. 2001. HMGN3a and HMGN3b, two protein isoforms with a tissue-specific expression pattern, expand the cellular repertoire of nucleosome-binding proteins. *J Biol Chem.* 276:25959–25969.
Wu J, Kim S, Kwak MS, Jeong JB, Min HJ, Yoon HG, Ahn JH, Shin JS. 2014. High mobility group nucleosomal binding domain 2 (HMGN2) SUMOylation by the SUMO E3 ligase PIAS1 decreases the binding affinity to nucleosome core particles. *J Biol Chem.* 289:20000–20011.
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.
Zhou BR, Feng H, Kato H, Dai L, Yang Y, Zhou Y, Bai Y. 2013. Structural insights into the histone H1-nucleosome complex. *Proc Natl Acad Sci U S A.* 110:19390–19395.
Zhu N, Hansen U. 2010. Transcriptional regulation by HMGN proteins. *Biochim Biophys Acta.* 1799:74–79.