Purifying Selection and Birth-and-death Evolution in the Histone H4 Gene Family

Helen Piontkivska, Alejandro P. Rooney, and Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University

Histones are small basic proteins encoded by a multigene family and are responsible for the nucleosomal organization of chromatin in eukaryotes. Because of the high degree of protein sequence conservation, it is generally believed that histone genes are subject to concerted evolution. However, purifying selection can also generate a high degree of sequence homogeneity. In this study, we examined the long-term evolution of histone H4 genes to determine whether concerted evolution or purifying selection was the major factor for maintaining sequence homogeneity. We analyzed the proportion (p_S) of synonymous nucleotide differences between the H4 genes from 59 species of fungi, plants, animals, and protists and found that p_S is generally very high and often close to the saturation level (p_S ranging from 0.3 to 0.6) even though protein sequences are virtually identical for all H4 genes. A small proportion of genes showed a low level of p_S values, but this appeared to be caused by recent gene duplication. Our findings suggest that the members of this gene family evolve according to the birth-and-death model of evolution under strong purifying selection. Using histone-like genes in archaebacteria as outgroups, we also showed that H1, H2A, H2B, H3, and H4 histone genes in eukaryotes form separate clusters and that these classes of genes diverged nearly at the same time, before the eukaryotic kingdoms diverged.

Introduction

Histones are highly conserved eukaryotic proteins and are responsible for the packaging of chromosomal DNA into nucleosomes (Pereira et al. 1997; Sandman and Reeve 2000). The eukaryotic nucleosome contains approximately 200 bp of DNA wrapped around an octamer of four different classes of histones (H2A, H2B, H3, and H4) and is loosely associated with a linker histone (H1). Among prokaryotes, one lineage of archaebacteria, *Euryarchaeota*, is known to have histone-like proteins. These histone-like proteins also bind and wrap DNA into nucleosomal structures (Bailey et al. 2000), but their relatively small size and absence of diversification into separate classes clearly distinguishes them from eukaryotic histones (Wu et al. 2000). In general, histone protein sequences are highly conserved within and between species. However, protists show a moderate amount of protein sequence diversity compared with other eukaryotes (Sadler and Brunk 1992; Bernhard and Schlegel 1998). This higher level of histone sequence diversity in protists may be attributed to the fact that they do not have condensed chromatin (e.g., *Trypanosoma* spp.; Espinoza et al. 1996), unlike higher eukaryotes (i.e., fungi, plants, and animals).

Among higher eukaryotic species, H4 proteins are highly conserved and show almost identical amino acid sequences within and between species (fig. 1). Although purifying selection certainly plays an important role in maintaining the high level of H4 protein sequence conservation, the observed amino acid sequence homogeneity is often explained by concerted evolution (Dover 1982; Maxson et al. 1983; Taylor, Wellman, and Marzluff 1986; Matsuo and Yamazaki 1989; DeBry and Marzluff 1994; Wang et al. 1996*a*, 1996*b;* Baldo, Les, and Strausbaugh 1999; Liao 1999). Concerted evolution can be defined as a process whereby individual members of a gene family do not evolve independently but instead evolve together as a unit by means of gene conversion or unequal crossing-over (Smith 1974; Arnheim 1983). In general, concerted evolution is expected to generate a higher degree of sequence similarity among multiple copies of genes within species than between species. However, histone H4 protein sequences are very similar even between distantly related species, such as animals and plants. This suggests that the major force for H4 protein homogeneity is purifying selection at the protein level.

Nevertheless, concerted evolution may take place at the DNA level. If this is the case, the DNA sequences of different member genes will be very similar within species but different between species. In the presence of strong purifying selection without concerted evolution, DNA sequence differences will be observed primarily at the synonymous sites. In this case, if the member genes evolve independently following the model of birth-and-death evolution as proposed by Nei and Hughes (1992), the DNA sequences of different member genes can be very different both within and between species at synonymous sites (Nei, Gu, and Sitnikova 1997). The birth-and-death model of evolution assumes that new genes are created by repeated gene duplication and that some of the duplicate genes are maintained in the genome for a long time, whereas others are deleted or become nonfunctional (Nei and Hughes 1992; Ota and Nei 1994; Nei, Gu, and Sitnikova 1997). It is possible to distinguish between the hypotheses of concerted evolution and birth-and-death evolution under strong purifying selection by examining the extent of synonymous differences within and between species. If intra-

Abbreviations: RD, replication dependent; RI, replication independent.

Key words: histone H4, concerted evolution, birth-and-death evolution, purifying selection.

Address for correspondence and reprints: Helen Piontkivska, Institute of Molecular Evolutionary Genetics, Pennsylvania State University, 328 Mueller Lab, University Park, Pennsylvania 16802. E-mail: oxp108@psu.edu.

Present address: Department of Biological Sciences, P.O. Box GY, Mississippi State University.

Mol. Biol. Evol. 19(5):689–697. 2002

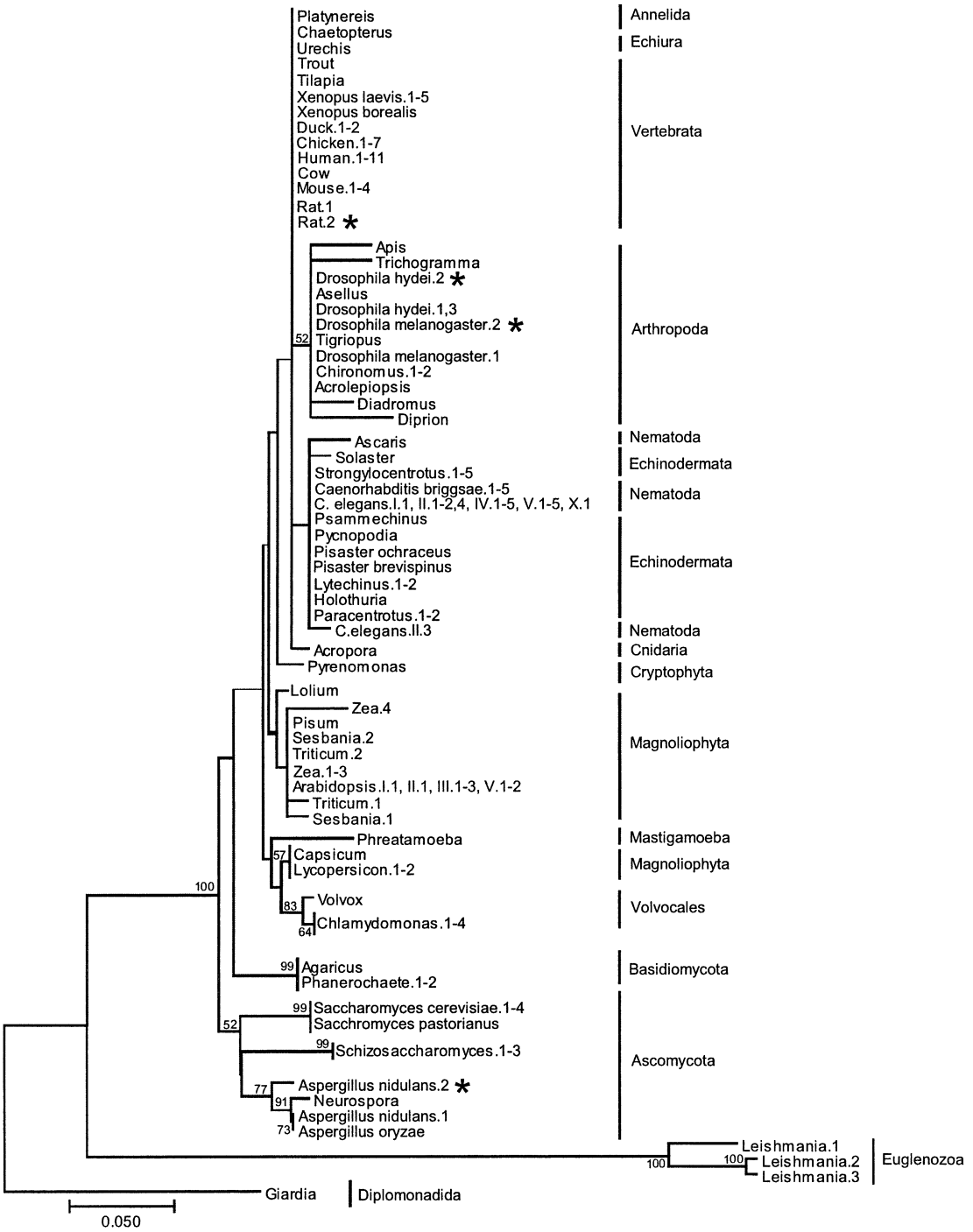

FIG. 1.—Phylogenetic tree of histone H4 amino acid sequences. Uncorrected *p*-distance was used. Complete-deletion option was used. Bootstrap values are based on 1,000 replications, and only those greater than 50% are shown. H4 sequence of *G. lamblia* was used as outgroup. RI sequences are marked by asterisks (*). See http://mep.bio.psu.edu/databases for sequence identification.

specific synonymous differences were nearly as high as interspecific differences, this would suggest that birth-and-death evolution under strong purifying selection is the dominant force. In this paper we investigate the mode of evolution of the histone H4 gene family by examining the proportion of synonymous (p_S) and nonsynonymous (p_N) differences per site between histone H4 genes both within and between species. The results obtained are presented subsequently.

Materials and Methods

A total of 137 histone H4 gene sequences from 4 protists, 11 plants, 8 fungi, and 36 animal species were obtained from GenBank. The complete genome sequence of *Caenorhabditis elegans* is now available (*C. elegans* Sequencing Consortium, 1998), and we included the entire set of histone H4 gene sequences from this species, i.e., 16 gene sequences. In addition, partial se-

quence data from the genome of *Caenorhabditis briggsae*, which is currently being sequenced, are publicly available. We have therefore included five H4 genes sequenced from this species. The numbers of gene sequences used from the other species are presented as follows (gene numbers are shown in brackets). Protists: *Giardia* (1), *Leishmania* (3), *Phreatamoeba* (1), and *Pyrenomonas* (1). Fungi: *Agaricus* (1), *Aspergillus nidulans* (2), *Aspergillus oryzae* (1), *Neurospora* (1), *Phanerochaete* (2), *Saccharomyces pastorianus* (1), *Saccharomyces cerevisiae* (4), and *Schizosaccharomyces* (3). Plants: *Chlamydomonas* (4), *Volvox* (1), *Arabidopsis* (7), *Capsicum* (1), *Lolium* (1), *Lycopersicon* (2), *Pisum* (1), *Sesbania* (2), *Solanum* (1), *Triticum* (2), and *Zea* (4). Animals: *Acropora* (2), *Ascaris* (1), *Chaetopterus* (1), *Platynereis* (1), *Urechis* (1), *Asellus* (1), *Tigriopus* (1), *Acrolepiopsis* (1), *Apis* (1), *Chironomus* (2), *Diadromus* (1), *Diprion* (1), *Drosophila hydei* (3), *Drosophila melanogaster* (2), *Trichogramma* (1), *Holothuria* (1), *Lytechinus* (2), *Paracentrotus* (2), *Pisaster brevispinus* (1), *Pisaster ochraceus* (1), *Psammechinus* (1), *Pycnopodia* (1), *Solaster* (1), *Strongylocentrotus* (5), *Tilapia* (1), trout (1), *Xenopus borealis* (1), *Xenopus laevis* (5), chicken (7), cow (1), duck (2), human (11), mouse (4), and rat (2). Both mRNA and genomic DNA sequences were analyzed. Nucleotide sequences were aligned by using Clustal X (Thompson et al. 1997) and checked visually for any possible errors afterwards. We have also used histone-like gene sequences from *Euryarchaeota* to study the origin of histone classes H1, H2A, H2B, H3, and H4 in eukaryotes. The representative sequences of eukaryotic histones from human, Drosophila, Arabidopsis, yeast, and Giardia were used for this purpose. A list of species names, sequence identifications, and GenBank accession numbers are available at http://mep.bio.psu.edu/databases.

The extent of nucleotide divergence was estimated by using the uncorrected *p* distance (Nei and Kumar 2000). The proportions of synonymous (p_S) and nonsynonymous (p_N) differences per site were computed by the modified Nei-Gojobori method (Zhang, Rosenberg, and Nei 1998). Phylogenetic trees were constructed by the neighbor-joining (NJ) method (Saitou and Nei 1987). All analyses were conducted by using the computer program MEGA, Version 2.1 (Kumar et al. 2001). The H4 gene of *Giardia lamblia* was used to root the tree for eukaryotic genes, as the *Giardia* lineage is believed to be the first to diverge from all other eukaryotes (Roger et al. 1998; Wu et al. 2000).

Results

Amino Acid Sequence Divergence

Histone proteins are classified into two different types based on their expression patterns: replication-dependent (RD) and replication-independent (RI) histones. RI histones are expressed continuously throughout the cell cycle, whereas RD histones are expressed only during the S-phase. However, H4 is the only histone protein that does not generally exhibit amino acid differences between RD and RI histones (Grimes et al. 1987; Akhmanova, Miedema, and Hennig 1996). The only species known to have distinct RI and RD protein sequences is the fungus *Aspergillus nidulans*, where RI proteins differ by two amino acids from RD proteins. Interestingly, the RD H4 gene of *Aspergillus* has an intron in the same position as the RD H4 gene of *Neurospora*. This suggests that these genes are more closely related to each other than either one is to the H4 RI gene from *Aspergillus*. Similarly, RI and RD H4 genes appear to have separate evolutionary histories in species that possess identical RD and RI H4 protein sequences.

In most cases, there are only very few amino acid differences between H4 protein sequences of different species, even if the species are highly divergent (fig. 1). For example, the H4 proteins from humans and the annelid worm *Platynereis dumerilii* show identical amino acid sequences (fig. 1), even though these two species diverged almost 800 MYA (Nei, Xu, and Glazko 2001). However, H4 proteins from the protist species used in this study display an unusually high level of sequence divergence (fig. 1). In this case, the majority of the variable sites are concentrated in the amino- and carboxyl-terminal regions of the protein. This relatively high level of divergence is not surprising if we note that the chromatin of these protist species does not condense during cell division (Aslund et al. 1994; Espinoza et al. 1996). Thus, purifying selection appears to be somewhat relaxed in the histone proteins of these species.

Nucleotide Sequence Divergence

The phylogeny of H4 genes based on nucleotide sequences is shown in figure 2. The extent of overall nucleotide sequence divergence is substantially higher than that of protein sequence divergence. However, because H4 proteins show little sequence variation, p_N is very low for most sequence comparisons (tables 1 and 2). Consequently, most of the nucleotide sequence variation is in the form of synonymous substitution. The phylogeny presented in figure 2 shows that the genes from the same species do not necessarily cluster together. However, different clusters of the phylogenetic tree are weakly supported by the bootstrap test. This again suggests that the genes from a species are no more closely related to each other than they are to genes from a different species. For example, in human and *Arabidopsis*, the majority of intraspecific p_S values are as high as the p_S values between animal, plant, and fungi species. In contrast, p_N values are very small even between different eukaryotic kingdoms (table 1).

In general, the extent of synonymous differences was very high both within and between species. The mean p_S ($\bar{p}_S$) commonly ranged from 0.2 to 0.6 between genes from the same species (table 2). In fact, synonymous substitutions have apparently reached the saturation level in many species, as shown by high p_S values (0.4–0.74) for individual pairwise comparisons (Nei, Rogozin, and Piontkivska 2000). Interestingly, the range of p_S values was nearly the same for both within- and between-species comparisons. For example, $\bar{p}_S$ ranged from 0.3 to 0.6 in all vertebrates, except in chicken and

FIG. 2.—Phylogenetic tree of representative H4 coding sequences. The number of synonymous nucleotide differences per site was used. Pairwise-deletion option was used. Bootstrap values are based on 1,000 replications, and only those greater than 50% are shown. H4 sequence of *G. lamblia* was used as outgroup.

Table 1
Numbers of Synonymous p_s (Below Diagonal) and Nonsynonymous p_N (Above Diagonal) Differences Per Site (×100) in H4 Genes of *Aspergillus* (Asp), Human (H), and *Arabidopsis* (A)

Gene	Asp. n.1	Asp. n.2	Asp.o.	H.1	H.2	H.3	H.4	H.5	A.1	A.2	A.3	A.4	A.5
Asp.n.1 . .		1	0	7	5	6	6	5	10	12	10	10	10
Asp.n.2 . .	31		1	7	5	6	6	5	10	12	10	10	10
Asp.o. . . .	22	40		7	5	6	6	5	10	12	10	10	10
H.1.	59	63	59		2	1	1	1	6	6	6	5	5
H.2.	56	56	63	49		1	1	0	7	8	6	6	7
H.3.	56	57	53	53	49		1	0	7	8	6	5	5
H.4.	67	68	65	59	53	59		1	6	8	6	6	7
H.5.	59	63	61	58	56	55	63		6	8	6	5	6
A.1.	52	63	52	65	71	72	74	81		2	3	4	4
A.2.	78	79	73	78	84	81	80	86	56		3	4	4
A.3.	74	74	74	78	82	74	75	78	63	66		3	3
A.4.	65	68	63	78	82	75	78	86	52	58	58		0
A.5.	64	69	62	81	82	75	79	85	54	56	61	7	

NOTE.—Asp.n. = *Aspergillus nidulans*, Asp.o. = *A. oryzae; Arabidopsis* genes from chromosomes I (1), II (2), and III (3–5) are presented.

duck (table 2). In fact, $\bar{p}_S$ was sometimes higher within species than between species. For example, $\bar{p}_S$ between human and mouse genes is smaller than $\bar{p}_S$ for intrahuman comparisons (0.480 and 0.542, respectively). Two frog species, *X. laevis* and *X. borealis*, which diverged about 15–20 MYA (Knochel et al. 1986), exhibit interspecific clustering of putatively orthologous genes (fig. 2 and table 2).

There are, however, some genes that show a low degree of intraspecific divergence at synonymous sites. For example, chicken and duck genes demonstrated low levels of H4 nucleotide sequence divergence. However, the level of divergence of the 5′- and 3′-flanking regions is rather high for the majority of these genes, considering the overall coding sequence identity (table 3). This and the fact that two pairs of chicken genes show a very high level of sequence similarity, even though they are located on the opposite DNA strands and separated by other histone genes, suggest that a recent gene duplication or gene conversion occurred, followed by an inversion to produce this high level of sequence similarity. However, the high sequence similarity in chicken and duck can also be attributed to an extremely high GC content in these genes, because the GC content at the third codon position is 95%–97%. A similar situation was observed in chicken and duck histone H3 genes (Rooney, Piontkivska and Nei 2002). Therefore, we cannot exclude the possibility of gene conversion in these genes, but at this point it is difficult to distinguish between the hypotheses of concerted evolution and birth-

Table 2
Numbers of Synonymous p_s (SE) (×100) and Nonsynonymous p_N (SE) (×100) Differences Per Site in H4 Genes from Vertebrate and Echinoderm Species

	p_s (SE)	p_N (SE)[a]
Intraspecific comparisons		
Chicken	1.6 (1.2)	0
Duck	1.1 (1.1)	0
Xenopus laevis (genes 1–4)	27.4 (3.1)	0.3 (0.3)**[b]
Human	54.2 (1.5)	0.7 (0.2)**
Mouse	25.8 (2.9)	0.4 (0.3)**
Rat	54.3 (4.5)	0**
Lytechinus	13.8 (4.3)	0*
Paracentrotus	33.5 (5.3)	0**
Strongylocentrotus	36.7 (3.3)	1.3 (0.6)**
Interspecific comparisons		
Chicken versus Duck	10.5 (2.9)	0.5 (0.5)**
X. laevis. 5 versus *X. borealis*	29.0 (6.6)	0**
X. laevis. 1–4 versus *X. borealis.*	41.9 (5.0)	0.5 (0.5)**
Mammals (average)	50.1 (5.5)	0.4 (0.3)**
Trout versus *Tilapia*	32.4 (6.1)	0.5 (0.4)**
Pisaster brevispinus versus *Pisaster ochraceus*	2.3 (1.4)	0

[a] There was no amino acid replacement associated with p_N values greater than zero, but the synonymous substitutions have occurred at the sixfold degenerate site coding for arginine (for example, AGA vs. CGG). Such changes were considered nonsynonymous in this study.

[b] In all *Z*-test pairwise comparisons $p_s > p_N$; however, the asterisks mark the significance level of the comparison (* indicates $P < 0.01$, ** indicates $P < 0.001$). SE indicates the standard error computed by the bootstrap method.

Table 3
Pairwise Comparisons of 5′ (Below Diagonal) and 3′ (Above Diagonal) Untranslated Regions of Chicken Histone H4 Genes Based on *p*-distance

	1	2	3	4	5	6	7
1		0.559	0.589	0.429	0.448	0.554	0.488
2[a]	0.573		0.023	0.408	0.403	0.562	0.476
3[a]	0.608	0.354		0.437	0.435	0.567	0.480
4	0.431	0.407	0.407		0.098	0.415	0.378
5	0.442	0.446	0.459	0.028		0.435	0.379
6[b]	0.563	0.530	0.546	0.400	0.428		0.293
7[b]	0.641	0.629	0.610	0.400	0.421	0.150	

[a] and [b] indicate chicken gene pairs with reverse orientation.

and-death evolution in these genes. To settle the problem, a complete genome analysis of chicken and duck genes as well as those from other avian species would be necessary.

Origin and Evolution of Histones H1, H2A, H2B, H3, and H4 in Eukaryotes

Eukaryotic genomes appear to be a chimera of eubacterial and archaebacterial genomes (Rivera et al. 1998), and histone genes seem to be descendants of ancient archaebacterial genes because no histone-like genes were found in eubacterial species (Pereira et al. 1997; Bailey et al. 2000). In eukaryotes, there are five classes of histones, i.e., H1, H2A, H2B, H3, and H4;

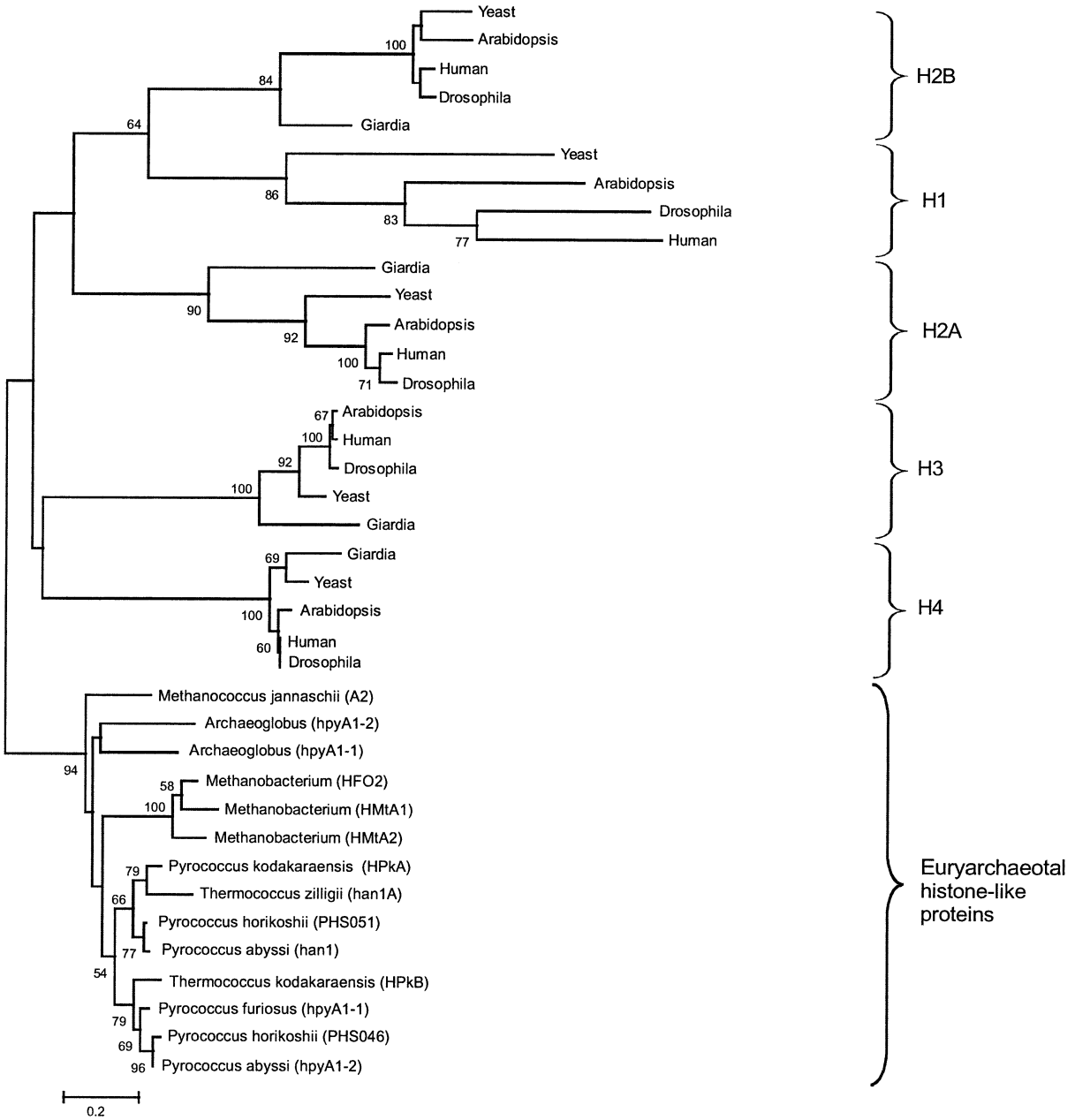

FIG. 3.—Phylogenetic tree of representatives from eukaryotic and euryarchaeotal histone–like proteins built using Poisson correction distance. Brackets mark histone-like proteins of *Euryarchaeota* and five major classes of eukaryotic histones. For *Euryarchaeota* protein names are also indicated.

because most of the eukaryotic genomes contain all of these histone genes, they must have evolved before the diversification of eukaryotic kingdoms. However, the origin and the divergence of these genes are largely unknown. We therefore conducted a phylogenetic analysis of these genes using archaebacterial genes as outgroups. The phylogenetic tree obtained by using Poisson correction (PC) distance for amino acid sequences is presented in figure 3. This tree shows that the five classes of histone genes are all monophyletic. However, the bootstrap support for the interior branches separating these five groups is rather low. This occurred partly because the number of amino acids used was small, but it is also possible that all five types of histone genes evolved nearly simultaneously, though the linker histone H1 might have arisen a little later. The latter hypothesis is also supported by the fact that the genome of one of the most primitive eukaryotes, *Giardia*, seems to be lacking H1 sequence (McArthur et al. 2000; Wu et al. 2000), though this can also be attributed to a gene loss event. Figure 3 also shows that histones H4 and H3 are the most conserved among all histone protein classes, whereas histone H1 is least conserved. This can be attributed to the crucial role of H3 and H4 proteins because the $(H3 + H4)_2$ tetramers initiate the nucleosome assembly and comprise the structural homologs of archaeal nucleosome (Sandman, Pereira, and Reeve 1998).

Discussion

Our results suggest that the amino acid sequence similarity among histone H4 genes is maintained primarily by strong purifying selection rather than by concerted evolution. At the nucleotide level, the numbers of synonymous differences between member genes from the same species are generally very large and often near the saturation level. This high level of synonymous differences suggests that H4 genes are subject to birth-and-death evolution at the DNA level and that many genes have persisted in the genome for a long time. This is quite interesting, considering the fact that H4 proteins from distantly related species (e.g., human, trout, and chicken) are identical (fig. 1). This long-term conservation of protein sequences can only be explained by strong purifying selection.

If H4 genes evolve according to the model of birth-and-death evolution under strong purifying selection, pseudogenes may be generated (Nei and Hughes 1991; Nei, Gu, and Sitnikova 1997). Indeed, H4 pseudogenes have been found in *X. laevis* (Turner et al. 1983), mice (Liu, Liu, and Marzluff 1987; DeBry 1998), humans (Kardalinou et al. 1993; Albig and Doenecke 1997), and Arabidopsis (Tacchini and Walbot 1995). Our analysis of *C. elegans* genome has suggested that there is at least one H4 pseudogene. Some (i.e., *Arabidopsis* pseudogene) of these pseudogenes appear to have emerged quite recently, whereas others (e.g., human and *C. elegans* pseudogenes) seem to be quite old, as shown by the level of sequence divergence from other genes (table 4).

Table 4
Human, *Caenorhabditis elegans* and *Arabidopsis* H4 Pseudogene Divergence, Based on *p*-distance (SE)

Pseudogene	Divergence Versus Functional Genes	Average Within-Species Divergence of Functional Genes
Human .1	0.368 (0.035)	0.163 (0.014)*
Human .2	0.366 (0.029)	0.163 (0.014)*
C. elegans	0.559 (0.028)	0.089 (0.010)*
Arabidopsis	0.134 (0.017)	0.148 (0.013)

* Marks the significance level of *Z*-test comparison between pseudogene divergence versus functional gene divergence ($P < 0.001$); SE indicates standard error computed by the bootstrap method.

We have shown that members of the histone H4 gene family do not evolve in a concerted manner in long-term evolution. Similar findings have also been reported in the highly conserved histone H3 family (Rooney, Piontkivska, and Nei 2002) and the ubiquitin gene family (Nei, Rogozin, and Piontkivska 2000). Furthermore, the model of birth-and-death evolution applies to many immune system gene families, such as the MHC (Nei and Hughes 1992; Gu and Nei 1999), immunoglobulin (Ota and Nei 1994), TCR (Su and Nei 2001), and ribonuclease genes (Zhang, Dyer, and Rosenberg 2000), as well as other multigene families (Duda and Palumbi 2000; Robertson 2000). It appears that birth-and-death evolution is the major mode of evolution of multigene families in eukaryotes.

Acknowledgments

This work was supported by grants from NIH (GM20293) and NASA (NCC2-1057) to M.N.

LITERATURE CITED

AKHMANOVA, A., K. MIEDEMA, and W. HENNIG. 1996. Identification and characterization of the *Drosophila* histone H4 replacement gene. FEBS Lett. **388**:219–222.

ALBIG, W., and D. DOENECKE. 1997. The human histone gene cluster at the D6S105 locus. Hum. Genet. **101**:284–294.

ARNHEIM, N. 1983. Concerted evolution of multigene families. Pp. 38–61 *in* M. NEI and R. K. KOEHN, eds. Evolution of genes and proteins. Sinauer Associates, Sunderland, Mass.

ASLUND, L., L. CARLSSON, J. HENRIKSSON, M. RYDAKER, G. C. TORO, N. GALANTI, and U. PETTERSSON. 1994. A gene family encoding heterogeneous histone H1 proteins in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. **65**:317–330.

BAILEY, K., S. PEREIRA, J. WIDOM, and J. REEVE. 2000. Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution. J. Mol. Biol. **303**:25–34.

BALDO, A. M., D. H. LES, and L. D. STRAUSBAUGH. 1999. Potentials and limitations of histone repeat sequences for phylogenetic reconstruction of *Sophophora*. Mol. Biol. Evol. **16**:1511–1520.

BERNHARD, D., and M. SCHLEGEL. 1998. Evolution of histone H4 and H3 genes in different ciliate lineages. J. Mol. Evol. **46**:344–354.

C. ELEGANS SEQUENCING CONSORTIUM. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282**:2012–2018.

DEBRY, R. W. 1998. Comparative analysis of evolution in a rodent histone H2A pseudogene. J. Mol. Evol. **46**:355–360.

DEBRY, R. W., and W. F. MARZLUFF. 1994. Selection on silent sites in the rodent H3 histone gene family. Genetics **138**: 191–202.

DOVER, G. 1982. Molecular drive: a cohesive mode of species evolution. Nature **299**:111–117.

DUDA, T. F. JR., and S. R. PALUMBI. 2000. Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. Mol. Biol. Evol. **17**:1286–1293.

ESPINOZA, I., G. C. TORO, U. HELLMAN, and N. GALANTI. 1996. Histone H1 and core histones in Leishmania and Crithidia: comparison with *Trypanosoma*. Exp. Cell Res. **224**: 1–7.

GRIMES, S., P. WEISZ-CARRINGTON, H. DAUM, J. SMITH, L. GREEN, K. WRIGHT, G. STEIN, and J. STEIN. 1987. A rat histone H4 gene closely associated with the testis-specific H1t gene. Exp. Cell Res. **173**:534–545.

GU, X., and M. NEI. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. Mol. Biol. Evol. **16**:147–156.

KARDALINOU, E., S. EICK, W. ALBIG, and D. DOENECKE. 1993. Association of a human H1 histone gene with an H2A pseudogene and genes encoding H2B.1 and H3.1 histones. J. Cell. Biochem. **52**:375–383.

KNOCHEL, W., E. KORGE, A. BASNER, and W. MEYERHOF. 1986. Globin evolution in the genus *Xenopus*: comparative analysis of cDNAs coding for adult globin polypeptides of *Xenopus borealis* and *Xenopus tropicalis*. J. Mol. Evol. **23**: 211–223.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, and M. NEI. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17**:1244–1245.

LIAO, D. 1999. Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. **64**:24–30.

LIU, T. J., L. LIU, and W. F. MARZLUFF. 1987. Mouse histone H2A and H2B genes: four functional genes and a pseudogene undergoing gene conversion with a closely linked functional gene. Nucleic Acids Res. **15**:3023–3039.

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

MAXSON, R., R. COHN, L. KEDES, and T. MOHUN. 1983. Expression and organization of histone genes. Annu. Rev. Genet. **17**:239–277.

MCARTHUR, A. G., H. G. MORRISON, J. E. NIXON et al. (15 co-authors). 2000. The *Giardia* genome project database. FEMS Microbiol. Lett. **189**:271–273.

NEI, M., X. GU, and T. SITNIKOVA. 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., and A. L. HUGHES. 1991. Polymorphism and evolution of the major histocompatibility complex loci in mammals. Pp. 222–247 *in* R. SELANDER, A. CLARK, and T. WHITTAM, eds. Evolution at the molecular level. Sinauer Associates Inc., Sunderland, Mass.

———. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. Pp. 27–38 *in* K. TSUJI, M. AIZAWA, and T. SASAZUKI, eds. 11th histocompatibility workshop and conference. Oxford University Press, Oxford.

NEI, M., and S. KUMAR. 2000. Molecular evolution and phylogenetics. Oxford University Press, Oxford.

NEI, M., I. B. ROGOZIN, and H. PIONTKIVSKA. 2000. Purifying selection and birth-and-death evolution in the ubiquitin gene family. Proc. Natl. Acad. Sci. USA **97**:10866–10871.

NEI, M., P. XU, and G. GLAZKO. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc. Natl. Acad. Sci. USA **98**:2497–2502.

OTA, T., and M. NEI. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. Mol. Biol. Evol. **11**:469–482.

PEREIRA, S. L., R. A. GRAYLING, R. LURZ, and J. N. REEVE. 1997. Archaeal nucleosomes. Proc. Natl. Acad. Sci. USA **94**:12633–12637.

RIVERA, M. C., R. JAIN, J. E. MOORE, and J. A. LAKE. 1998. Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA **95**:6239–6244.

ROBERTSON, H. M. 2000. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. Genome Res. **10**:192–203.

ROGER, A. J., S. G. SVARD, J. TOVAR, C. G. CLARK, M. W. SMITH, F. D. GILLIN, and M. L. SOGIN. 1998. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc. Natl. Acad. Sci. USA **95**:229–234.

ROONEY, A. P., H. PIONTKIVSKA, and M. NEI. 2002. Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. Mol. Biol. Evol. **19**:68–75.

SADLER, L. A., and C. F. BRUNK. 1992. Phylogenetic relationships and unusual diversity in histone H4 proteins within the *Tetrahymena pyriformis* complex. Mol. Biol. Evol. **9**: 70–84.

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

SANDMAN, K., S. L. PEREIRA, and J. N. REEVE. 1998. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. Cell. Mol. Life Sci. **54**:1350–1364.

SANDMAN, K., and J. N. REEVE. 2000. Structure and functional relationships of archaeal and eukaryal histones and nucleosomes. Arch. Microbiol. **173**:165–169.

SMITH, G. P. 1974. Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. **38**:507–513.

SU, C., and M. NEI. 2001. Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. Mol. Biol. Evol. **18**:503–513.

TACCHINI, P., and W. WALBOT. 1995. Nucleotide sequence of a histone H4 pseudogene in *Arabidopsis thaliana*. Plant Physiol. **108**:1747.

TAYLOR, J. D., S. E. WELLMAN, and W. F. MARZLUFF. 1986. Sequences of four mouse histone H3 genes: implications for evolution of mouse histone genes. J. Mol. Evol. **23**:242–249.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**: 4876–4882.

TURNER, P. C., T. C. ALDRIDGE, H. R. WOODLAND, and R. W. OLD. 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**:4093–4107.

WANG, Z. F., T. KRASIKOV, M. R. FREY, J. WANG, A. G. MATERA, and W. F. MARZLUFF. 1996*a*. Characterization of the

mouse histone gene cluster on chromosome 13: 45 histone genes in three patches spread over 1 Mb. Genome Res. **6**: 688–701.

WANG, Z. F., R. TISOVEC, R. W. DEBRY, M. R. FREY, A. G. MATERA, and W. F. MARZLUFF. 1996*b*. Characterization of the 55-kb mouse histone gene cluster on chromosome 3. Genome Res. **6**:702–714.

WU, G., A. G. MCARTHUR, A. FISER, A. SALI, M. L. SOGIN, and M. MULLER. 2000. Core histones of the amitochondriate protist, *Giardia lamblia*. Mol. Biol. Evol. **17**:1156–1163.

ZHANG, J., K. D. DYER, and H. F. ROSENBERG. 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.

ZHANG, J., H. F. ROSENBERG, and M. NEI. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

FUMIO TAJIMA, reviewing editor

Accepted January 2, 2002