Independent evolutionary origin of histone H3.3-like variants of animals and *Tetrahymena*

Thomas H.Thatcher[+], Jennifer MacGaffey, Josephine Bowen, Stuart Horowitz[§], Donald L.Shapiro and Martin A.Gorovsky*
Department of Biology, University of Rochester, Rochester, NY 14627, USA

ABSTRACT

All three genes encoding histone H3 proteins were cloned and sequenced from *Tetrahymena thermophila*. Two of these genes encode a major H3 protein identical to that of *T.pyriformis* and 87% identical to the major H3 of vertebrates. The third gene encodes hv2, a quantitatively minor replication independent (replacement) variant. The sequence of hv2 is only 85% identical to the animal replacement variant H3.3 and is the most divergent H3 replacement variant described. Phylogenetic analysis of 73 H3 protein sequences suggests that hv2, H3.3, and the plant replacement variant H3.III evolved independently, and that H3.3 is not the ancestral H3 gene, as was previously suggested (Wells, D., Bains, W., and Kedes, L. 1986. J. Mol. Evol., 23: 224–241). These results suggest it is the replication independence and not the particular protein sequence that is important in the function of H3 replacement variants.

INTRODUCTION

The synthesis of most histones is closely linked to DNA replication in the cell cycle (1–4). However, some histones are constitutively synthesized, even in non-dividing cells. They are quantitatively minor non-allelic histone primary sequence variants called basal or replacement variants. They are distinguished from replication dependent histones (5–8) because they are synthesized and deposited in the nuclei of non S phase (G_1, G_2, quiescent) cells, replacing their normal counterparts in nucleosomes.

Replacement variants for histone H3 have been found in plants (9), mammals (10), birds (11), *Drosophila* (12), and *Tetrahymena* (13,14). The genes encoding the vertebrate H3 variants (H3.3) are notably different from major H3 genes in that they contain introns and encode polyadenylated messages. The proteins encoded by these vertebrate H3 genes differ in slight but similar ways from the major replication dependent H3s. Because histones evolve slowly, it is not clear whether the small differences in protein sequence between replication and replacement H3s are

indicative of functional variation at the protein level or are simply neutral polymorphisms. In the latter case, a conserved function requiring histone turnover, unrelated to DNA replication, might be performed by any H3 protein whose synthesis is uncoupled from replication. An analysis of codon use and intron position among H3 genes (15) led to the suggestion that the replacement variant H3.3 is actually the ancestral H3 gene, and that replication coupled H3 genes evolved from H3.3, losing their introns and polyadenylation signals in the process. However, this analysis included only H3.3 genes from vertebrates, all of which encode the same protein. The more recent finding that *Drosophila* contains an H3.3 gene encoding a protein whose sequence is identical to that of vertebrates (12) argues that this gene diverged from the major H3 gene(s) early in the evolution of multicellular eukaryotes and is consistent with (but does not prove) the suggestion that H3.3 is the ancestral gene.

Tetrahymena and other ciliated protozoans contain two nuclei: mitotically dividing, transcriptionally inert micronuclei and amitotically dividing, transcriptionally active somatic macronuclei. The histones and histone genes of *Tetrahymena thermophila* are similar to those of higher eukaryotes in a number of fundamental ways (16). *Tetrahymena* chromatin is arranged in typical periodic and particulate nucleosomes (17,18) containing the four core histones and about 200 bp of DNA. Most if not all of the secondary modifications found on core histones in higher eukaryotes are also found on *Tetrahymena* core histones (16). In addition, macronuclei contain two core histone variants called hv1 and hv2, which are present in sub-stoichiometric amounts relative to the major core histones and are absent from micronuclei (13). Histone hv1 is an H2A variant which is conserved in evolution and appears to be preferentially associated with transcriptionally active chromatin (13,19). hv2 is a replacement variant of H3; like the H3.3 variants of vertebrates it is synthesized and deposited in nuclei of non-growing cells (20). Since the sequences of *Tetrahymena* histones are among the most divergent described (16), a study of hv2 should shed light on the evolution and function of H3 replacement variants. Similar amino acid changes in the *Tetrahymena* and animal H3

---

*To whom correspondence should be addressed

Present addresses: [+]School of Medicine and Dentistry, University of Rochester, 601 Elmwood Avenue, PO Box 363, Rochester, NY 14642 and [§]CardioPulmonary Research Institute, Suite 503, Winthrop University Hospital, 222 Station Plaza North, Mineola, NY 11501, USA

replacement variants would strongly suggest a requirement for important structural features in the proteins themselves and an early evolutionary origin for the H3.3 histone subclass. On the other hand, a lack of conserved amino acid replacements would argue that most eukaryotes simply require an H3 gene whose expression is unlinked to DNA replication.

Southern blots of macronuclear DNA and northern blots probed with a yeast histone H3 gene indicated *Tetrahymena* has three H3 genes and three different size classes of polyadenylated H3 mRNA (20). With the yeast probes we previously isolated two genomic fragments which encode the amino-terminal halves of two H3 genes, H3-I and H3-II (21). These fragments were used to probe a cDNA library made from RNA from starved *Tetrahymena*. Such a library should be highly enriched for hv2 clones since only a single large, polyadenylated H3 message is detected in starved cells which deposit only newly synthesized hv2 into macronuclei (20). We report here the isolation and sequencing of a *Tetrahymena* H3 cDNA clone encoding hv2 from that library and its corresponding genomic clone, as well as the isolation and sequencing of genomic clones encoding the entire H3-I and H3-II genes. A phylogenetic tree of H3 peptide sequences suggests that, in contrast to H2A variants, which have a common early origin, H3 replacement variants appear to have arisen independently, at least twice, as relatively recent events.

MATERIALS AND METHODS

Cell culture and isolation of mRNA

Tetrahymena thermophila strain CU428 (a gift from Peter Bruns, Cornell University) were grown to log phase in enriched proteose peptone and starved in 10 mM Tris, pH 7.4, as previously described (22). Total RNA was isolated essentially as described (23). Polyadenylated RNA was isolated using Hybond messenger affinity paper (Amersham, Arlington Heights, IL) according to the manufacturer's instructions.

Colony and plaque lifts, Southern blots, and DNA hybridizations

Phage plaques and plasmid colonies were lifted onto nitrocellulose filters as described (24). Oligonucleotide probes were labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase (New England Biolabs, Beverly, MA) according to the enzyme supplier's recommendations. DNA probes were labelled by the random primer method (25). Hybridizations (Southern blots and plaque or colony lifts) were carried out in 30% formamide, 0.5 M NaCl, 5 mM Tris−Cl, pH 7.5, 0.01% sodium pyrophosphate, 0.1% SDS, 0.01 mM EDTA, 1×Denhardt's solution (26), and 0.1 mg/ml sheared and denatured herring sperm DNA (Sigma, St Louis, MO) at 60°C. Blots were washed at 60°C in 2×SSPE (26), 0.1% SDS. When oligonucleotides were used as probes, the tetramethyl ammonium chloride hybridization procedure of Wood *et al.* was used (27).

Construction of a starved cDNA library and isolation of a histone hv2 cDNA clone

RNA from starved cells was used to construct a cDNA library in λgt11 as described (19). Plaque lifts were screened with the 4 kb Eco RI insert of pTt999.1 (21) which contains a complete copy of a *T.thermophila* H4 gene (H4-II) and a portion of an H3 gene (H3-II) encoding the amino-terminal half of the major H3 protein. Positive plaques were picked for secondary screening, plated, and lifted in duplicate. To discriminate between clones

containing H3 genes and H4 genes, one set of filters was hybridized with pTt999.1 (H3 + H4) as described above, and the other with a yeast H3 gene (28). Two H3 clones were found in this library, one of which, λTtS101, was used for this study. A 470 bp insert was subcloned into M13mp18 and mp19 and single stranded DNA was sequenced by the dideoxy method (29) using the BRL (Gaithersburg, MD) sequencing kit according to manufacturer's instructions.

Cloning and sequencing of the hv2 gene HHT3 from genomic DNA

The sequence of λTtS101 was compared to that of the amino terminal halves of the major H3 genes, H3-I and H3-II, which had previously been cloned (21). A 31 base oligonucleotide was constructed from a region of hv2 that differed from both of the major H3 genes (indicated on Fig. 2). Labelled oligonucleotide was hybridized to a Southern blot of macronuclear DNA digested with various restriction enzymes. A 1566 bp band from an Eco RI−Hind III double digest which hybridized to the oligonucleotide was cloned into pGemini 1 vector (Promega, Madison, WI) using the size selected library method described in Stargell and Gorovsky (23). Taq I restriction fragments of this insert were subcloned into M13mp18 and mp19 and sequenced as described above.

Cloning and sequencing of the major H3 genes HHT1 and HHT2 from genomic DNA

Previously, clones containing portions of the coding regions of H3-I and H3-II were isolated from a *T.thermophila* Eco RI genomic library (21). These H3-I and H3-II clones contain the first 78 codons encoding the major histone H3 protein and 4.3 kb and 3.7 kb of upstream sequence respectively. The H3-II clone also contains the entire coding sequence of the histone H4-II, which is transcribed divergently beginning 340 bp upstream of the ATG of H3-II (Fig. 1). A 127 bp Bgl II−Taq I fragment of the H3-II 5′ flanking region, and a 872 bp Acc I−Bgl II fragment of the H3-I 5′ flanking region were used as gene specific probes for restriction mapping of macronuclear DNA. These analyses suggested that both genes reside on separate Bgl II−Cla I genomic DNA fragments of 2.4 kb for H3-I and 2.0 kb for H3-II (data not shown). A size selected Bgl II−Cla I plasmid library was constructed (23) and recombinant colonies were identified by hybridization to the gene-specific probes. The H3-I and H3-II genes were sequenced (29) progressively using Sequenase (USB, Cleveland, OH) according to manufacturer's instructions with oligonucleotide primers corresponding to previously sequenced regions.

Phylogenetic analysis of histone protein sequences

Histone protein sequences were compiled from the Protein Identification Resource database (PIR, release 33.0, June, 1992) and by translating DNA sequences obtained from GenBank (release 77.0, June, 1993). The histone compilations of Wells (30,31) were used as a guide. Only complete protein sequences were used. Where both protein and DNA sequences were available for the same organism, the translation of the DNA sequence was used, as it is unambiguous, compared to protein sequences that may have been derived by purification from mixtures of primary sequence variants. We elected not to conduct the analysis on the DNA sequences encoding these histones for three reasons. Many organisms, including *Tetrahymena* (32) and yeast (33), are known to use codons in a highly biased fashion

that is independent of the encoded protein. In a DNA-based analysis of sequence relatedness, these codon biases would tend to decrease the apparent similarity between sequences in organisms with different codon preferences. Second, over the long time since the divergence of the organisms included in the analysis, it is likely that synonymous second and third positions have been saturated with mutations and hence have been randomized, except to the extent influenced by codon bias. Any analysis of non-synonymous substitutions is essentially an analysis of the encoded proteins. Finally, some histones have only been sequenced directly, and the coding sequences are unavailable.

The sequences were aligned using the PileUp program in the GCG Sequence Analysis package (34), which uses a simplified form of the progressive sequence alignment method (35). Evolutionary distances (d) were calculated from the proportion of amino acid identity (S) with a Poisson correction, $d = -\ln S$. These values were used to construct phylogenetic trees by the neighbor-joining method (36). Parsimony analysis was carried out using the program PAUP version 3.1.1 (37) on subsets of protein sequences chosen from major phylogenetic lineages determined by neighbor-joining methods (the complete sequence sets are too large to analyze by parsimony methods).

The histone protein alignments are available from the authors by electronic mail at marty@mag.biology.rochester.edu, or on computer diskette in Macintosh or PC format; the compiled protein sequences are available in Macintosh format only. Please send a blank disk and your complete mailing address and phone number with your request.

RESULTS

The complete sequence of the *Tetrahymena* H3 gene family

We have determined the sequence of three *Tetrahymena thermophila* histone H3 genes (Fig. 2). All of the bands detected by a yeast H3 probe on a Southern blot of macronuclear DNA (20) can be accounted for by the restriction maps of these clones (Fig. 1); thus we have cloned the entire complement of *T.thermophila* H3 genes.

HHT1 and HHT2 encode the major H3 protein, a 135 amino acid protein that is identical to the major H3 of *Tetrahymena pyriformis* (14) and 87% identical to the major animal H3 (Fig. 3). The DNA sequence of HHT2 is 93% identical to HHT1 over the coding region. The genes contain 3 and 4 TAA codons, which encode glutamine in *Tetrahymena* (20), respectively; neither gene contains an intron.

HHT3 encodes an H3 protein having 16 replacements relative to the major H3 but only three differences relative to the quantitatively minor H3(2) of *T.pyriformis* (Fig. 3). At the DNA level, HHT3 is 86% identical to HHT1 within the coding region and contains 3 TAA codons; it does not contain any introns. The absence of introns in the H3 genes is not reflective of a general lack of introns in histone or other Pol II genes of *Tetrahymena*. Of 35 *T.thermophila* genes available in GenBank as of April, 1993, 11 contain introns (data not shown), including the genes for the histones H1 (38), and hv1 (39).

Several lines of evidence indicate that HHT3 encodes the H3 replacement variant hv2. hv2 is the only H3 deposited in macronuclei of starved cells and only one of the three H3 messages is present in starved cells (20); HHT3 was originally cloned as a cDNA from a starved cell library. Additionally, a Sac I–Hind III fragment from the 3' nontranscribed region of HHT3 hybridizes only to the starved cell specific H3 message (data not shown).

hv2 is functionally analogous to other H3 replacement variants but is structurally dissimilar from them

hv2 is a replication-independent histone H3 variant similar in its regulation to animal H3.3. and plant H3.III. As for vertebrate H3.3s, hv2 is the only H3 protein to be synthesized and deposited in the nuclei of non-dividing cells. hv2 is encoded by the longest of the H3 mRNAs, which has a 3' untranslated region at least 515 nt long (data not shown), longer than any *Tetrahymena* histone message yet determined (40,41). These characteristics are reminiscent of vertebrate H3.3s, which are also encoded on messages containing long 3' untranslated regions (42,43).

Despite these similarities, the amino acid sequence of hv2 is substantially different from the animal and plant H3 variants (Fig. 3). The H3.3 proteins from 5 vertebrates and *Drosophila* (AH3.3) are completely identical and differ from the major animal H3 (AH3) at only 4 positions—31, 87, 89, and 90. Plant H3.I (PH3.I) and H3.III (PH3.III) also differ from one another at four positions, including 31, 87, and 89. These three positions are also sites where *T.thermophila* hv2 and *T.pyriformis* H3(2) differ from the major *Tetrahymena* H3 (TetH3) (Fig. 3, boxed). Although positions 31, 87, and 89 appear to vary in the H3 replacement variants, none of the substitutions at these positions are conserved. An alanine at position 31 in the major (replication dependent) H3s is replaced with valine in *Tetrahymena* hv2 and H3(2), with serine in AH3.3, and with threonine in PH3.III. A serine at position 87 in the replication H3s is changed to glutamine in hv2 and H3(2), alanine in AH3.3 and histidine in PH3.III,
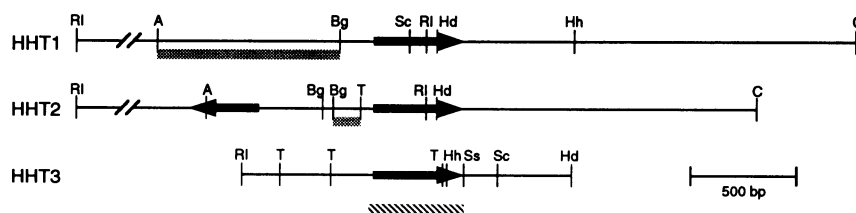
Figure 1. Genomic organization of the *T.thermophila* H3 genes. Portions of HHT1 and HHT2 were previously cloned as Eco RI fragments containing the 5' half of the coding regions. For this report HHT1 and HHT2 were cloned as Bgl II–Cla I fragments containing the entire coding sequence. Black arrows indicate transcription units. The leftward-pointing arrow with HHT2 is the gene HHF2, encoding histone H4, which is divergently transcribed from HHT2. Shaded bars in HHT1 and HHT2 indicate flanking sequence-specific probes used to clone the genes. The striped bar in HHT3 shows the location of the cDNA clone λTtS101. Restriction enzyme sites are abbreviated A, Afl II; Bg, Bgl II; C, Cla I; Hh, Hha I; Hd, Hind III; RI, Eco RI; Sc, Sca I; Ss, Sst I; T, Taq I.

```
HHT3                                                        -640 GAATTCTTGACTTATGAATAAACAAATTAACCTGTAGGTC  -601
HHT3  TAGTCTAATCTAGTCTGGTATGGTCTTGTTTATTTGATATGGTGGGGGTGGGGAGAAACAAATATCAAAAAAAATAAAGAAAGTGGAAGTTAGTTAGTTATTTATTTAGGTGTTTT  -481
HHT3  GTTTTTTAGCTGATATATTTTTTTTAGTTTGTTATCGAGTAATGAGGAGTACAATCTAACTAGCAAACTAACTTAGTAAATAAAGAAATTAATTAATTAGTATTCAGTAAATAAGGATGAG  -361
HHT3  GGAGATGCGAGGGAGGTTGGGGTTGAGAAATTGGTAGTTGGATGGATGAATGATGGATGAGGTAGGTAAATGAGAATGGTGGTTAATAAAAATATACATACAGCTTGATTACTATGGAAA  -241

HHT1                                                    -172  GATCTGGATCATTTTGAAGAATTGTCCAATCAGATTGAAGATTGCGTTTAAG  -121
HHT2                       -211  ATCTCAAATTTGATTGGATAAAATCTCAAAAATCTGATTTCCTTTCAAATCATCCAATCAAAAATTGTTCATATCTTCTCTTTCCAGGATTA
HHT3  GAAACAAGAATATTAATTATCGACAGCACTTTAAGTATATATTTTTGATTAACTTACGTAATTGGAATAGAAATGGTGATTCCTTTGCTATAAACACTCTTTTTTTAGAAAATAAGAATA

HHT1  ATTGAATTTTTGAAATTTATTTAAGAGAGGCAAGAGAATGAAAATTTGTATAATTACTTAATTTAAAAGAAAAATTAATTTTAAAAAAGATAAACTCATTCATTATACAAACTAATAAAA  -1
HHT2  TGCCAAACATTTCAAATGATCTTGCCCTAATTGTTGTCGATAAAGAATTTTATTAAAAAAAAATAATAACCTTCAAAAGAAAAGTCAAAAGACAATCCACTATAAATACATAAGCAAAA
HHT3  TAATATTTTTCAAAACAAAACAAAATTAAATTAAAAAAAAAATTTAATCAAAAAAAAAAACCTCAAAGTTTTTTGAAAAAAGCAATTAAAAAAATAATTTTCACACAAAACAGTAACAAAA

HHT1  ATGGCTAGAACTAAATAAACTGCTAGAAAATCAACTGGTGCTAAGGCTCCCAGAAAGCAACTCGCTTCCAAGGCCGCCAGAAAGTCTGCCCCCGCCACCGGTGGTATTAAGAAGCCCCAC  120
HHT2  ATG              G  C          C      AT            C                    T        C
HHT3  ATG              G       T     T     A     A       T                 T  T T     G  C              T

HHT1  AGATTCAGACCTGGTACCGTCGCCTTGAGAGAAATCAGAAAGTACTAAAAGTCCACTGATTTACTTATCAGAAAGCTCCCCTTCCAAAGATTAGTTAGAGATATTGCTCACGAATTCAAG  240
HHT2              TC  C           C             GT  G                      G  C
HHT3  A                       T     A T   C     T             C T  C          ATG  A  G

HHT1  GCTGAACTCAGATTCTAATCTTCTGCCGTTCTTGCTCTCCAAGAAGCTGCTGAAGCTTACCTCGTTGGTCTCTTCGAAGACACCAACCTCTGCGCTATCCACGCCAGAAGAGTCACCATC  360
HHT2                                                    C             T       T  G           T          T  T  T
HHT3  AG   TA     C   CCAA   TA C     C          C       C        C  T A      T                         T       T  T  T

HHT1  ATGACCAAGGATATGCAACTCGCCAGAAGAATCAGAGGTGAAAGATTCTGAGGAGTGTTTTGACAAAGATTCTTTTCAAACAATCGAGTTAAATTAATATTAAATAATTGCTGAATAATC  480
HHT2                                            T       T       TGAGCATAATATAACAACTAGTCTCTAAATAATCAACATACCACATATATATCATTACTTAATTTATCAAAA
HHT3              C  C  C  T  T       T     A       TGAGAGCTCTCTCTAATAATTAACTTAATATATAATACACACATATAATATAATGTTTTCTTATTATGGGAA

HHT1  AATCGAGTTTGTAGTTTAAGATTTATGAAGAGTTGGCTTTGATAGCAAGCATGTTTTGGAATCAAGCATAGAAATGTTAATAATAATAATAACAAACTAAATAATTAGCAAAGTAATTTG  600
HHT2  TATCATATGTATATTATAACTTTCGTGTGAGTACATAAAACCTTTTTTAGAGTTTTCGAGTTTTCAAAAAAGATAATCCAAAATAACTTCCTCCATATATATAAAATATGCATATATATAA
HHT3  TGCAGCTTGCATCTTGATGCTTTTTACTCTCTTTGATTATAAATAAACTTCATTGATTCATACGGATACATAAAATCATACTCTATATCAAGTACTTTATTCACTCATTCATTGATGAGT

HHT1  ATCTGATGAAAATAAAAATTAATAAAATAATCATATAAAAATGCATATAAATTTGACTACTTAATTAAACAATATTAGATTAAATGTATTTCAATTTTGTTATATATATTTTTCATTTCT  720
HHT2  AAATAATTAATTAAATTCCACAATATAAAATTTATTTGTTTTTTTTATTCAATATCATTCATTTGTATTTATTACTAAAAAGGTTTCTATCTTTAGCTAACTTATCCATGCAAAATTTATGA
HHT3  GAGTGAATGAATGAAAAAACATGGGTGGTTGATTAACTTATTAATTAATTTAATATAAAGACTGCATATATGCCAATCAACATGCATACATGCATGAAATGTAATTAATTGGTCCTGACA

HHT1  CGCTTCCATTAAATGCTTTTGAGAGGCTTCTCTCTCATAAATTTATAATCATTTTTCCAATTCTTTATTTCAAAAAACTTAATATATTTTATTTAAATTTCATTTACTTCCCCCTTTTTC  840
HHT2  AATTTAATTCAAATTATGCGAAAAAGCAAACAAACAAACAAACATATTTTGCAATTTAATAAGTTTTAATAAAATTACTTATTTTATAAATACAAATAGTGAGCATCACCATACTTTTAT
HHT3  AATAAATTATGTATAAATAAATAAATAAAACTATTAAAACGCATTGAGCCAAAAACAGCTTGCATACTGCATCTTCCAATTACTTGTTTCATACACTCTTATATAAATTTAAATTAGTGG

HHT1  CTTCTAACGATTCAGTATTTGTCGTCTATTGACTTCAAATCTTTAATTAAATTATCTTAATTAATACTGTTAAAACAAATAAATTAGAATCGCGCGCATTTTTCAGATTCAATATTTTAT  960
HHT2  TATTATATTTTATTTTATTTCCAATAAATAATGAATGAATGAATCTTGCAAATCCCTCCAAATCAAAATCAAAATTAATATTTATTTATATTTTTATTTAAAATTTACAACAATGTATCT
HHT3  TATCAAATCACATAAAATATATATAAATATATATTATGTATAATCTTAATTTTATTGCTAATCATCTCTTTCAATTGAAAAAGCTT  926

HHT1  AACATCTCTCAATTACCAAATCTAATTAAATTAATTTTCTTAAAAAAAATATTTTTCACCAATAATCATTTTTAGTACATTTATGAGATACTCATTTAACTTTATCAATACATTTTTCAT  1080
HHT2  TTATATTTTCTAAAATTTAAACTTTCAAATCCTAATTTAATATGTATTAAAAATTTTTATTGCTTACAACTAAATTTATTTAAATTAATATAAACAAGTTAATATTAAACTTTAG  1075

HHT1  TTAAATATTAAATTGATTATCATTTTTTTTT  1111
```
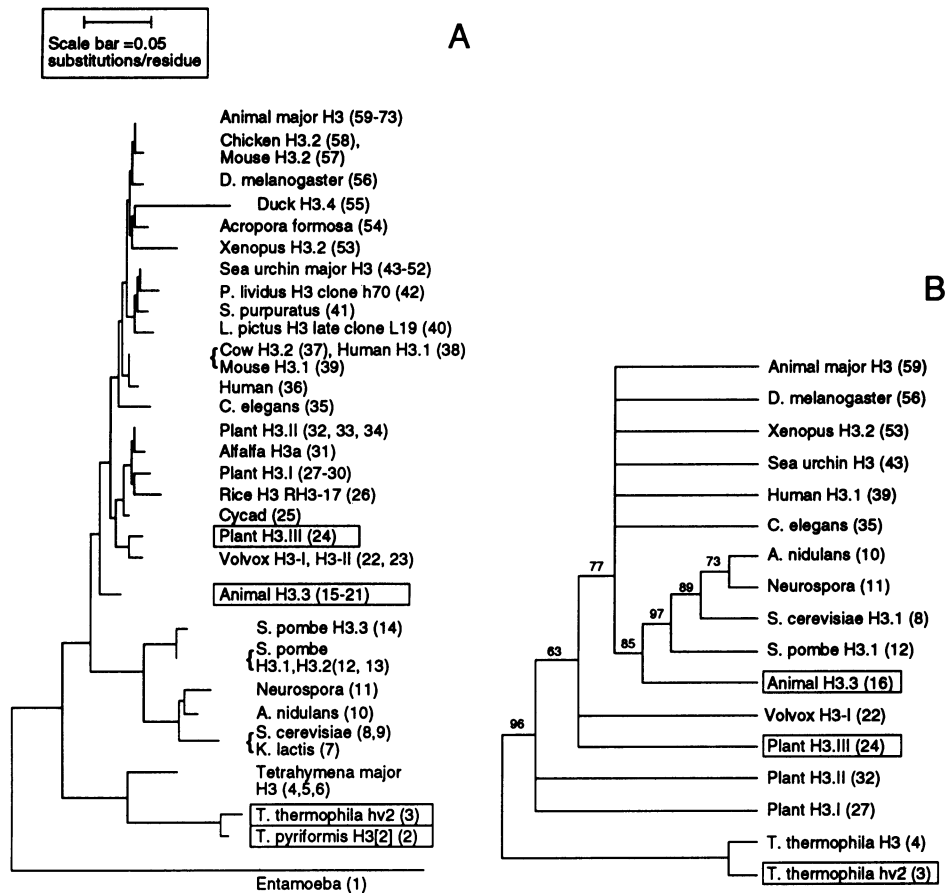
Figure 2. Sequence of the *T.thermophila* H3 genes. The sequences of HHT1, HHT2 and HHT3 are presented; within the coding region only differences are shown for HHT2 and HHT3. The underline in HHT3 indicates the sequence of the gene-specific oligonucleotide used to obtain the genomic copy of the gene. These sequences have been deposited in GenBank under accession numbers M87304, M87504, and M87305.

```
              1                                                                    70
   Tet H3     ARTKQTARKSTGAKAPRKQLASKAARKSAHATGGIKKPHRFRPGTVALREIRKYQKSTDLLIRKLPFQRL
   Tet hv2               V           T       VS   V     K                    T
   TP H3(2)              V           T       VS   V     K                    T
   AH3                   G           T            V     Y              R      E
   AH3.3                 G           T       S    V     Y              R      E
   PH3.I                 G           T            V                           E
   PH3.III               G           T       T    V     Y                     E

              71                                                                   135
   Tet H3     VRDIAHEFKAELRFQSSAVLALQEAAEAYLVGLFEDTNLCAIHARRVTIMTKDMQLARRIRGERF
   Tet hv2         M  M  SDI    Q  I                                   LH
   TP H3(2)        M  M  NDI    Q  I
   AH3        E    QD    TD        M      S                    K     P     I          A
   AH3.3      E    QD    TD    A  IG      S                    K     P     I          A
   PH3.I      E    QD    TD        A                           K     P     I          A
   PH3.III    E    QD    TD    H                               K     P     I          A
```

Figure 3. Comparison of the sequences of major and variant H3 peptides from *Tetrahymena*, animals, and plants. The complete sequence of the *Tetrahymena* major H3 protein is shown (TetH3) with differences for *T.thermophila* hv2 (Tet hv2), *T.pyriformis* H3(2) (TP H3(2)), Mouse H3.2 (AH3), Mouse H3.3 (AH3.3), *Arabidopsis* H3.I (PH3.I), and *Arabidopsis* H3.III (PH3.III). Boxed residues at positions 31, 87, and 89 are highly conserved in replication-dependent major H3s.

while the valine at residue 89 is replaced with isoleucine in the *Tetrahymena* and animal variants but not replaced in PH3.III. In addition to these three sites, TetH3 and hv2 differ at 13 other positions, 7 of which occur at residues where both TetH3 and hv2 differ from AH3 and PH3.I.

H3 replacement variants arose at least twice in evolution

Although hv2 is clearly an H3 replacement variant by its expression, its sequence is distinct from those of other H3 replacement variants. To analyze the evolution of the H3 variants,

a phylogenetic tree was constructed using the neighbor-joining (NJ) method of Saitou and Nei (36) on 73 H3 sequences compiled from the GenBank and PIR databases (see Materials and Methods). The total evolutionary divergence between two sequences is represented by the sum of the horizontal branch lengths between the two sequences. Vertical distances are for illustration purposes only.

The NJ tree for H3 (Fig. 4A) indicates that the three types of H3 replacement variants represented (animal H3.3, plant H3.III and *Tetrahymena* hv2 and H3(2)) are not monophyletic;

Figure 4. Phylogenetic trees of histone H3 proteins. **A.** Neighbor-joining tree drawn from 73 aligned H3 protein sequences. The scale bar indicates the branch length that corresponds to 0.05 substitutions per position (the alignment contains 137 positions). Numbers in parentheses are keys to the GenBank accession numbers (see below). Boxed sequences indicate replication-independent H3 variants. **B.** Parsimony tree drawn from a subset of 17 H3 protein sequences using the heuristic algorithm of PAUP. Numbers along the branches are bootstrap confidence values for 250 replicates; branches with values smaller than 50 are not shown. Plant H3 nomenclature is according to (52). GenBank and PIR accession numbers are given below (PIR indicated by *). Tetrahymena major H3 includes *T. thermophila* HHT1 and HHT2, M87304 and M87504, and *T. pyriformis* H3, A28852*. Plant H3.III is *Arabidopsis* H3, X60429. Plant H3.II includes Alfalfa, X13673; Pea, A02631*; Wheat, A26014*. Plant H3.I includes *Arabidopsis*, M35387; Maize, M36658; Parsley, M77493; Rice, X13678. Animal H3.3 includes Chicken H3.3A, M11392, M11667; Chicken H3.3B, M11393; Clam, M17876; *D.melanogaster* H3.3q, X53822; Human H3.3, X05855, X05856, X05857; Mouse H3.3, X13605; Rabbit H3-3A, X51897. Sea urchin major H3 includes *D.imbricatus*, X07505; *L.pictus* late H3, X00593; *L.pictus* early H3, X00628; *P.brevispinus*, X54112; *P.helianthoides*, X54114; *P.lividus*, M25281; *P.miliaris*, V01140; *P.ochraceus*, X54113; *S.drobachiensis*, M36921; *S.stimpsoni*, X54115. Animal major H3 includes Buffalofish, A02627*; Cat shark, A02626*; Chicken, X02218; Chicken H3-IV, X62291; Chicken H3-V, X62292; *Chironomus thummi*, X56335; Cow H3.1, A02624*; *D.hydei*, X17072; Duck H3.1, X14732; Mouse H3.2, M33989; *Platyneris*, X53330; Trout, X01064; *Tigriopus californicus*, X52393; *Urechis caupo*, X58895; *Xenopus laevis* H3.1, A24279*. Others are 1, L02418; 2, B28852*; 3, M87305; 7, X14230; 8, X00724; 9, X00725; 10, X55548; 11, X01612; 12, X05222; 13, X05223; 14, X05224; 22, X06963; 23, X06964; 25, A23604*; 26, M15664; 31, M35867; 35, X15634; 36, M26150; 37, A02624*; 38, X57128; 39, M32460; 40, J01175; 41, X03952; 42, M36919; 53, J00984; 54, L11067; 55, M14396; 56, X14215; 57, X16148; 58, M61155.

that is, they do not have a single common ancestor. The H3 to which the *Tetrahymena* variants hv2 and H3(2) are most closely related is the *Tetrahymena* major H3. Animal H3.3s branch off the tree prior to the divergence of plants and animals, while plant H3.III is found within the monophyletic plant branch. Although this suggests that plant H3.III may have evolved from an earlier H3.3-type gene, other NJ trees which do not contain the *A.nidulans* sequence have a different topology, in which animal H3.3s and fungal H3s are monophyletic and branch after the divergence of plant H3s (data not shown). To investigate this discrepancy, a maximum parsimony analysis was carried out on a subset of H3 proteins from major phylogenetic lineages with the program PAUP (37) (Fig. 4B). The use of a subset containing fewer taxa permits a more rigorous analysis using parsimony methods. The parsimony tree places the divergence of animals

H3.3s and fungal H3s two branches deeper than plant H3.III with a high degree of confidence. A NJ tree of the same subset was identical to one of the minimum length parsimony trees (data not shown). This confirms our NJ analyses that suggested that animal H3.3s evolved after the plant-animal divergence and were closely related to animal and fungal H3s. We carried out additional parsimony and NJ analyses on several different subsets of H3 proteins and found that most NJ and all parsimony phylogenies placed animal H3.3s and fungal H3s on the same branch, after the divergence of plants (data not shown). Thus, the parsimony method appears to be more robust with respect to this data set, and suggests that animal, plant and *Tetrahymena* H3 replacement variants arose independently. Both NJ and parsimony analyses agree unequivocally that *Tetrahymena* replacement variant arose independently of other H3 variants.

Table 1. New designations for *Tetrahymena thermophila* histone genes

Protein	New gene name	Old gene name	GenBank accession no.	Reference
H1	HHO	H1	M14854	(38)
Micronuclear linker histone	MLH	MLH	M87306	(53)
H2A.1	HTA1	*	L18892	X.Liu and M.A.Gorovsky, unpublished observations
H2A.2	HTA2	*	L18893	X.Liu and M.A.Gorovsky, unpublished observations
hv1 (H2A variant)	HTA3	hv1	X14137	(19)
H2B.1	HTB1	TH2B-1	M31332	(40)
H2B.2	HTB2	TH2B-2	M31333	(40)
H3	HHT1	H3-I	M87304	(21) and this report
H3	HHT2	H3-II	M87504	(21) and this report
hv2 (H3 variant)	HHT3	*	M87305	This report
H4	HHF1	H4-I	X04756	(54)
H4	HHF2	H4-II	X04755	(54)

* No previous gene designation.

Revised nomenclature for *Tetrahymena thermophila* histone genes

The histone gene complement of *Tetrahymena thermophila* consists of 12 genes encoding 10 histone proteins including an unusual H1 lacking the conserved central hydrophobic domain, a unique micronucleus specific linker histone that is totally unlike H1, and an H2A variant (hv1) that is homologous to mammalian H2A.Z and chicken H2A.F proteins (Table 1). However, as the sequences have been determined they have been given a variety of nonstandard names. In particular, the H3 genes were originally named H3-I and H3-II despite the fact that the partial sequences predicted identical peptides (21). Since, with this report the sequences of all 12 known *Tetrahymena* histone genes have been determined (10 by our laboratory), we propose standardizing the histone gene names. Thus, we have named the genes HHO (histone H-one), MLH (micronuclear linker histone), HTA (histone H-two-A), HHT (histone H-three), and so on. We have retained hv1 and hv2 as the names of the proteins encoded by HTA3 and HHT3 for historical reasons. Note that HHT1 and HHT2, and HHF1 and HHF2, encode identical major H3 and H4 proteins, respectively.

DISCUSSION

We have cloned and sequenced the three histone H3 genes from *Tetrahymena thermophila*. Thus, *T.thermophila* is one of the few eukaryotes for which the sequence of its entire histone gene family is known (see Table 1). The *T.thermophila* H3 genes encode the major replication-dependent H3 and the replication-independent (replacement) variant hv2, whose predicted amino acid sequence is very different from those of animal and plant H3 replacement variants.

Phylogenetic analyses of H3 protein sequences by the neighbor-joining and parsimony methods suggest that the three types of replacement variants arose independently at least two and probably three times in eukaryotic evolution. This uncertainty is due to a puzzling change in the topology of the NJ tree caused by the addition of the *A.nidulans* H3 sequence. We had constructed the H3 NJ tree several times with increasingly larger data sets and found the branch order highly resistant to the addition of new sequences and uniform in its placement of animal H3.3 and fungal H3s on the same branch (data not shown). We

do not know why the addition of the *Aspergillus* H3 changes the placement of H3.3, since the H3s from *Aspergillus* and *Neurospora* differ at only 4 positions and at 3 of those 4 positions the *Aspergillus* sequence is more like the other fungal H3s than *Neurospora* (data not shown). Given that animal H3.3 and fungal H3s were monophyletic for all parsimony analyses carried out on the entire H3 data set and on several different subsets, and that NJ trees constructed from the same subsets all placed H3.3 and fungal H3s on the same branch, we feel it is reasonable to conclude that H3.3 evolved after the branching of plants and animals, and thus independently of plant H3.III and *Tetrahymena* hv2.

Wells *et al.* (15) compared codon use among H3 and H3.3 genes (excluding plant H3s and *Tetrahymena* H3 and hv2, which were not available) and concluded that an H3.3 histone gene including introns and a polyadenylation signal was probably the ancestral H3 from which the current major H3s were derived. Our analysis differs from Wells' in several important respects. First, it includes variants from *Tetrahymena* and plants which were not previously available and which are highly informative. Additionally, we include as an outgroup an H3 from *Entamoeba*, an organism which rRNA phylogenies place as diverging from the main eukaryotic line prior to any of the other species included in the H3 tree (44). Finally, the H3 genes of some organisms, including *Tetrahymena* (32) and yeast (33), may have biased codon use that reflects the codon bias of the entire genome rather than any individual gene. Such codon bias is, of course, not reflected in an analysis of protein sequences.

Neither the NJ nor the parismony analysis rules out the possibility that an H3.3 gene is the common ancestor of animal and fungal H3s. In this scenario (a modification of Wells' hypothesis, see 15) a primitive animal−fungal ancestor contained a single type of H3 gene with introns and a polyadenylation signal. After the divergence of the fungi, animals duplicated this gene. One of the duplicates lost its polyadenylation signal and introns, acquired certain amino acid substitutions and DNA replication-coupled synthesis, and became the 'major' H3 gene of animals. The fungal gene retained its H3.3 type amino acid substitutions at positions 31, 87, and 89 while acquiring 11 to 12 additional substitutions. The current mode of regulation of fungal H3 genes is irrelevant to this model. However, the model suggests that any H3 genes which contain introns would have them at the same

place. Yet the introns of *Neurospora* and *Aspergillus* H3 are in different positions from each other and from those of chicken and human H3.3s (42,45−47).

Alternatively, fungal H3s may group with H3.3 in the phylogenetic analyses because they have converged on certain amino acid substitutions important for their function. Fungal H3s are cell-cycle regulated (47,48) but their expression in nondividing cells has not been investigated. If they are expressed constitutively as well, they might be expected to have amino acid substitutions typical of basal H3 variants. However, although the amino acids at positions 31, 87, and 89 (alanine, serine, and valine, respectively) are conserved among most replication dependent H3s and vary in basal variants and fungi, the three types of replacement variants have three different sets of amino acid substitutions at these positions. It is therefore impossible to associate specific amino acid substitutions with basal H3 variant-specific functions. The question of why these residues have become fixed in most replication-dependent H3s but are variable in replacement/fungal H3s is interesting but unanswered.

It can not be determined from sequence information alone whether animals, plants, fungi and *Tetrahymena* require a replacement H3 variant with certain structural features or simply any H3 synthesized basally. This question can be addressed by exchanging the coding regions of replication-dependent and independent H3 genes. Unfortunately, this experiment is impractical in higher eukaryotes where homologous recombination is rare and repeated histone genes and pseudogenes complicate matters. However, with the recent advent of gene disruption and replacement in *Tetrahymena* (49−51), which has a small and completely sequenced histone gene family, this approach should now be feasible.

ACKNOWLEDGEMENTS

REFERENCES

1. Busslinger,M., Schämperli,D. and Birnstiel,M.L. (1986) Cold Spring Harbor Symp. Quant. Biology, 51, 665−670.
2. Hentschel,C.C. and Birnstiel,M.L. (1981) Cell, 25, 301−313.
3. Schümperli,D. (1986) Cell, 45, 471−472.
4. Schümperli,D. (1988) Trends in Genetics, 4, 187−191.
5. Zweidler,A. (1984) In Stein,G.S., Stein,J.L. and Marzluff,W.F. (eds.), Histone Genes. John Wiley & Sons, New York, pp. 339−369.
6. Ferrari,N., Pfeffer,U., Profumo,A. and Vidali,G. (1992) Biochem. Int., 28, 239−248.
7. Waterborg,J.H. (1993) J. Biol. Chem., 268, 4912−4917.
8. Wunsch,A.M. and Lough,J. (1990) Cell Differ. Dev., 30, 19−25.
9. Chaubet,N., Clement,B. and Gigot,C. (1992) J. Mol. Biol., 225, 569−574.
10. Marzluff,W.F. and Graves,R.A. (1984) In Stein,G.S., Stein,J.L. and Marzluff,W.F. (eds.), Histone genes: Structure, organization, and regulation. John Wiley and Sons, New York, pp. 281−316.
11. Engel,J.D. (1984) In Stein,G.S., Stein,J.L. and Marzluff,W.F. (eds.), Histone genes: Structure, organization, and regulation. John Wiley and Sons, New York, pp. 263−280.
12. Fretzin,S., Allan,B.D., Van Daal,A. and Elgin,S.C.R. (1991) Gene, 107, 341−342.
13. Allis,C.D., Glover,C.V.D., Bowen,J.K. and Gorovsky,M.A. (1980) Cell, 20, 609−617.
14. Hayashi,T., Hayashi,H., Fusauchi,F. and Iwai,K. (1984) J. Biochem., 95, 1741−1749.
15. Wells,D., Bains,W. and Kedes,L. (1986) J. Mol. Evol., 23, 224−241.
16. Gorovsky,M.A. (1986) In Gall,J.G. (ed.), The Molecular Biology of Ciliated Protozoa. Academic Press, Inc., Orlando, pp. 227−261.
17. Gorovsky,M.A., Glover,C., Johmann,C.A., Keevert,J.B., Mathis,D.J. and Samuelson,M. (1978) Cold Spring Harbor Symp. Quant. Biology, XLII, 493−503.
18. Giri,C.P. and Gorovsky,M.A. (1980) Nucleic Acids Res., 8, 197−214.
19. White,E.M., Shapiro,D.L., Allis,C.D. and Gorovsky,M.A. (1988) Nucleic Acids Res., 16, 179−198.
20. Bannon,G.A., Calzone,F.J., Bowen,J.K., Allis,C.D. and Gorovsky,M.A. (1983) Nucleic Acids Res., 11, 3903−3917.
21. Horowitz,S. and Gorovsky,M.A. (1985) Proc. Natl. Acad. Sci. USA, 82, 2452−2455.
22. Calzone,F.J., Stathopoulos,V.A., Grass,D., Gorovsky,M.A. and Angerer,R.C. (1983) J. Biol. Chem., 258, 6899−6905.
23. Stargell,L.A. and Gorovsky,M.A. (1994) Mol. Cell. Biol., in press.
24. Benton,W.D. and Davis,R.W. (1977) Science, 196, 180−182.
25. Feinberg,A.P. and Vogelstein,B. (1983) Anal. Biochem., 132, 6−13.
26. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. Wood,W.I., Gitschier,J., Lasky,L.A. and Lawn,R.M. (1985) Proc. Natl. Acad. Sci. USA, 82, 1585−1588.
28. Smith,M.M. and Murray,K. (1983) J. Mol. Biol., 169, 641−661.
29. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463−5467.
30. Wells,D. and McBride,C. (1989) Nucleic Acids Res., 17 Suppl., r311−r346.
31. Wells,D. and Brown,D. (1991) Nucleic Acids Res., 19 Suppl., 2173−2188.
32. Martindale,D.W. (1989) J. Protozool., 36, 29−34.
33. Smith,M.M. (1984) In Stein,G.S., Stein,J.L. and Marzluff,W.F. (eds.), Histone Genes. John Wiley and Sons, New York, pp. 3−33.
34. Devereux,J., Haeberli,P. and Smithies,O. (1984) Nucleic Acids Res., 12, 387−395.
35. Feng,D-F. and Doolittle,R.F. (1987) J. Mol. Evol., 25, 351−360.
36. Saitou,N. and Nei,M. (1987) Mol. Biol. Evol., 4, 406−425.
37. Swofford,D.L. (1993) PAUP: Phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign, Illinois.
38. Wu,M., Allis,C.D., Richman,R., Cook,R.G. and Gorovsky,M.A. (1986) Proc. Natl. Acad. Sci. USA, 83, 8674−8678.
39. Van Daal,A., White,E.M., Elgin,S.C.R. and Gorovsky,M.A. (1990) J. Mol. Evol., 30, 449−455.
40. Nomoto,M., Imai,N., Saiga,H., Matsui,T. and Mita,T. (1987) Nucleic Acids Res., 15, 5681−5697.
41. Liu,X. and Gorovsky,M.A. (1993) Nucleic Acids Res., 21, 4954−4960.
42. Wells,D., Hoffman,D. and Kedes,L. (1987) Nucleic Acids Res., 15, 2871−2889.
43. Dodgson,J.B., Yamamoto,M. and Engel,J.D. (1987) Nucleic Acids Res., 15, 6294.
44. Sogin,M.L. (1989) Amer. Zool., 29, 487−499.
45. Woudt,L.P., Plastink,A., Kempers-Veenstra,A.E., Jansen,A.E.M., Mager,W.H. and Planta,R.J. (1983) Nucleic Acids Res., 11, 5347.
46. Brush,D., Dodgson,J.B., Choi,O-R., Stevens,P.W. and Engel,J.D. (1985) Mol. Cell. Biol., 5, 1307−1317.
47. Ehinger,A., Denison,S.H. and May,G.S. (1990) Mol. Gen. Genet., 222, 416−424.
48. Cross,S.L. and Smith,M.M. (1988) Mol. Cell. Biol., 8, 945−954.
49. Gaertig,J. and Gorovsky,M.A. (1992) Proc. Natl. Acad. Sci. USA, 89, 9196−9200.
50. Kahn,R.W., Andersen,B.H. and Brunk,C.F. (1993) Proc. Natl. Acad. Sci. USA, 90, 9295−9299.
51. Gaertig,J., Thatcher,T.H., Gu,L. and Gorovsky,M.A. (1994) Proc. Natl. Acad. Sci. USA, in press.
52. Waterborg,J.H. (1991) Plant Physiol., 96, 453−458.
53. Wu,M., Allis,C.D., Sweet,M.T., Cook,R.G., Thatcher,T.H. and Gorovsky,M.A. (1994) Mol. Cell. Biol., in press.
54. Horowitz,S., Bowen,J.K., Bannon,G.A. and Gorovsky,M.A. (1987) Nucleic Acids Res., 15, 141−160.