Common Features of Analogous Replacement Histone H3 Genes in Animals and Plants

Jakob H. Waterborg, Anthony J. Robertson*

Division of Cell Biology and Biophysics, School of Biological Sciences, University of Missouri–Kansas City, 5100 Rockhill Road, Kansas City, MO 64110-2499, USA

Abstract. Phylogenetic analysis of histone H3 protein sequences demonstrates the independent origin of the replacement histone H3 genes in animals and in plants. Multiple introns in the replacement histone H3 genes of animals in a pattern distinct from that in plant replacement H3 genes supports this conclusion. It is suggested that replacement H3 genes arose at the same time that, independently, multicellular forms of animals and of plants evolved. Judged by the degree of invariant and functionally constrained amino acid positions, histones H3 and H4, which form together the tetramer kernel of the nucleosome, have co-evolved with equal rates of sequence divergence. Residues 31 and 87 in histone H3 are the only residues that consistently changed across each gene duplication event that created functional replacement histone H3 variant forms. Once changed, these residues have remained invariant across divergent speciation. This suggests that they are required to allow replacement histone H3 to participate in the assembly of nucleosomes in non–S-phase cells. The abundant occurrence of polypyrimidine sequences in the introns of all replacement H3 genes, and the replacement of an intron by a polypyrimidine motif upstream of the alfalfa replacement H3 gene, suggests a function. It is speculated that they may contribute to the characteristic cell-cycle-independent pattern of replacement histone H3 genes by binding nucleosome-excluding proteins.

Introduction

Replacement histones have a cell-cycle-independent pattern of expression that continues when cell proliferation ceases (Van Holde 1989). This characteristically leads to a gradual increase in the relative abundance of these histone variant proteins in the chromatin of terminally differentiated cells and aging mature organisms (Zweidler 1984; Wells et al. 1989). Constitutively expressed animal histone variant H3.3 was identified as a functional replacement variant, replacing distinct H3.1 and H3.2 replication-dependent forms in chicken and mouse (Urban and Zweidler 1983; Zweidler 1984). Replacement histone H3 variants have been identified in alfalfa and *Arabidopsis* by their cell-cycle-independent pattern of synthesis and the high levels of acetylation and protein turnover that are indicative of the formation of replacement nucleosomes (Waterborg 1990, 1992; Chaubet et al. 1992; Kapros et al. 1992). Like their animal counterparts, these genes contain multiple introns (Chaubet et al. 1992; Robertson et al. 1995) with long untranslated flanking sequences (UTRs) in the mature transcripts that may function in mRNA stabilization (Marzluff 1989; Kapros et al. 1995).

Recently, phylogenetic analysis of histone H3 sequences has revealed that the histone H3.3-like replace-

---

* *Present address:* Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3280, USA
Correspondence to: J. H. Waterborg

ment histone of *Tetrahymena* (Thatcher et al. 1994) arose independently of the animal replacement H3.3 gene. This indicates that these replacement histones are used in functionally analogous processes. Study of the similarities between functionally analogous replacement histone H3 forms could reveal essential structural features of replacement histone H3 that confer the capability to assemble DNA into nucleosomes in nonreplicating cells.

The published phylogenetic analysis indicated that the plant replacement H3 histones might also be analogous rather than homologous to the animal replacement H3.3 histones (Thatcher et al. 1994). The analysis presented here confirms this suggestion, based on new sequences information on correction of database sequence errors that were identified by comparison with their published sources, and on comparison of gene clone and variant sequences within the same organism or among closely related species. Subsequently, this paper identifies common elements in the protein sequences and gene organization of the three analogous groups of replacement H3 genes that appear essential for the histone H3 replacement function. The possible relevance of these common elements is discussed in the context of the structure of the nucleosome (Van Holde 1989; Arents and Moudrianakis 1993) and in the light of new experimental observations.

Materials and Methods

Histone H3 and H4 protein sequences were collected from GenBank (release 88.0) and PIR (release 44.0) databases and are identified by their unique accession numbers. Swiss Protein Database (release 31.0) accession numbers are included if they provide information on histone sequences that was not available in GenBank or PIR files. Twenty-four distinct histone H3 protein sequences, numbered 1 to 24, and 21 distinct histone H4 protein sequences, numbered 25 to 45, are listed below with the descriptor tags used in tables and figures in bold. For each, accession numbers and species (binomen), trivial name, if significant, and applicable major and minor GenBank Taxonomy classifications are provided. The presence of a 3'UTR palindromic stem-loop structure is indicated as [Ω]. If a U7 snRNA-recognized downstream cleavage sequence is also present, this is marked as [Ω-U7]. In lower eukaryotic Chlorophyta and *Physarum* it is marked as (Ω-u?) because the presence of a U7-related sequence for these nonpolyadenylated transcripts has not been demonstrated (Wilhelm et al. 1988; Fabry et al. 1995). The presence of a putative polyadenylation signal in *Pyrenomonas* (Muller et al. 1994) is shown as [Ω-polyA]. Demonstrated mRNA polyadenylation in one or multiple species of a category is indicated by [polyA]. Protein sequence analysis without information on the 3'UTR of mature transcripts is given as [−]. The presence and location of introns [I =] is given for each species, where available from genomic and cDNA clones, as inside the 5'UTR of a transcript or by the first residue of the following exon, expressed as protein residue number, with the position within the codon between brackets. Residues 1 to 135 for histone H3 are numbered according to sequence 1, and residues 1 to 102 for histone H4 according to sequence 26. The positions of α-helices in the histone octamer model (Arents and Moudrianakis 1995) are marked by superimposed zigzags with the name of the helix. Inserted residues and unaligned sequence positions are counted in reference to the last preceding aligned residue. All currently accessible histone H3 and H4 sequences that provide or predict complete or significant parts of histone protein sequences are included. Single-pass automated sequences

are marked by EST and are typically not error-free. Recognized erroneous database entries are given with source for error correction. Demonstrated or suspected errors in sequencing or codon translation in sequences deposited are explicitly identified. Unless otherwise noted, these typically represent single-base read errors in single clones, absent in other analyses from the same organism, and are listed by residue number between {Ω}. Variability in single clones from single organisms, absent in multiple clones from closely related species, is noted similarly and generally excluded from analysis. Recognized pseudogenes are explicitly identified and excluded.

Phylogenetic relationships of histone protein sequences were determined by neighbor-joining analysis by MEGA version 1.02 (Kumar et al. 1993) by 1,000 bootstrap analyses with paired deletions. The results of raw data differences and calculation of p-, Poisson, and gamma distances were indistinguishable with bootstrap values within 3% for all methods, except for the branch points of highly divergent sequences 22 to 24 and 45. Maximum parsimony analysis was performed for animal and plant H3 histones with the *N. crassa* H3 as outgroup in 250 heuristic bootstrap analyses by the protein parsimony analysis program of PAUP (Swofford 1993). The complete H3 sequence set is too large to analyze by parsimony methods.

Histone H3 Sequences. **1. mammal** H3.1: [Ω-U7] Animalia, Mammalia *Homo sapiens,* man A02623, A40335, M26150, M60746, P16106, S28528, X00090, X57128, X83550, Z46261; *Bos taurus,* cow A02624, P16106; *Mus musculus,* mouse M32460, M32462, S06755, X01684, X16496; *M. pahari* X80324, X80325; *Rattus norvegicus* L19706. **2. animal** H3.2 and H3: [Ω-U7] Animalia, Mammalia *B. taurus* A02624, P16105; *M. musculus* JH0304, M32459, M32461, M33989, S06743, X01685, X16148 {112 codon translation error), X80328, Z30939; *M. pahari* X80326, X80327; *Cricetulus longicaudatus,* hamster M28265, X80330; Aves *Gallus gallus,* chicken A02625, J00869, JS0690, M61154, M61155, S17359, S18716, X02218, X62291, X62292; *Anas platyrhynchus,* duck X14732; Pisces *Salmo gairdneri,* trout B23220, X01064; *Ictiobus bubalus,* buffalo fish A02627; *Poroderma africanus,* catshark A02626; Amphibia *Xenopus laevis* A02634, {21, 26, 28, 86, 102 sequencing errors}, A24279, A93596, M21286, M21287, P02302, X03017, X03018, X03104, X72949, X72950; *Notophthalmus viridescens,* newt J00953; Tunicata *Styela plicata,* sea squirt JN0687, S64499 {88, 99, 127, 128, 131 codon translation and sequencing errors}; Echiura *Urechis caupo,* spoonworm X58895; Insecta *Drosophila melanogaster,* fruit fly A02630, P02299 {102 typing error}, X14215 {31 typing error}; *Drosophila hydei* S09655, X17072, X52576; *Drosophila virilis* M33982; *Chironomus thummi* S17992, S40438, X56335, X72803; *Leptothorax acervorum* X77742; Crustacea *Tigriopus californicus* M84797, X52393; Mollusca *Patella granatina,* sandpiper limpet B61286; Annelida *Platynereis dumerilii,* clamworm S11315, X53330. Excluded pseudogenes: *M. musculus* A02633, V00754; *Cairina moschata,* duck A23981, M14396, P06902. **3. Echinodermata** H3: [Ω-U7] Animalia, Echinodermata, Echinizoa *Psammechinus millaris* A02628, J01181, M10558, P02298, V01140, V01143, V01144; *Paracentrotus lividus* M25281, M36919, M36920; *Strongylocentrotus drobaciensis* M36921; *Strongylocentrotus purpuratus* A02629 {96, 102 codon translation errors}, J01173, P06352, V01356, X03952, {102 sequencing error}; *Lytechinus pictus* X00628; *Parechinus angulosus* A61286; Echinodermata, Asterozoa *Dermasterias imbricata* S01198, X07505; *Pisaster brevispinus* S01196, S20667, X07504, X54112; *Pisaster ochraceus* S01197, S20671, X07503, X54113; *Pycnopodia helianthoides* S20669, X54114; *Solaster stimpsoni* S20678, X54115. **4. animal-R** Replacement H3.3: [polyA] Animalia, Mammalia *H. sapiens* [I = 5'UTR, 42(3), 94(1)] A02622, A27501, M11353, M11354, P06351, X05855, X05856, X05857, Z48950; *M. musculus* B43805, S04186, X13605; *R. norvegicus* X73683; *Oryctolagus cuniculus,* rabbit S10168, X51897; Aves *G. gallus* [I = 5'UTR, 42(3), 94(1)] M11392, M11393, M11667, Y00392; Insecta *D. melanogaster* [I = 5'UTR, 94(1)] P33155 {120 typing error}, X53822, X81207, X82257; *D. hydei* [I = 5'UTR, 94(1)] X81206, X81208; *L. acervorum* X77740, X77741; Mollusca *Spisula*

solidissima, clam M17876. **5. Echinodermata-R** Replacement H3.3: [polyA] Animalia, Echinodermata *P. lividus* X76081. **6. Nematoda-1** major H3: [−] Animalia, Nematoda *Caenorhabditis elegans* P08898 (major protein variant). **7. Nematoda-2** minor H3: [Ω-U7] Animalia, Nematoda *C. elegans* X15634. Entries 6 and 7 do not fully represent histone H3 complexity in *C. elegans:* Z38112 (EST), (Vanfleteren et al. 1989). **8. Cnidaria** H3: [Ω-U7] Animalia, Cnidaria *Acropora formosa,* coral JQ0757, L11067, M60509, P22943, S67324. **9. algae** H3: [Ω-u?] Chlorophycota *Volvox carteri* S00940, X06963 [I = 46(2)], X06964 [I = 46(1)]; *Chlamydomonas reinhardtii* L41841 {90, 113 sequencing errors}, U16724, U16725, U16825. **10. plant-R** Replacement H3-III, H3.2 and H3.3: [polyA] Plantae, dicot *Arabidopsis thaliana* [I = 5'UTR, 24(2), 49(3), 80(1) (gene 1); 5'UTR, 24(2), 49(3) (gene 2)] S24346, X60429; *Medicago sativa,* alfalfa [I = 24(1), 49(3), 80(1)] B38309, S04521, U09458, U09460, U09461, U09462, U09463, U09464, U09465, X13676; *Lycopersicon esculentum,* tomato X83422; monocot *Hordeum vulgare,* barley A02632, P06353; *Zea mays,* corn T14800 (EST), T23317 (EST), T23357 (EST); *Lolium temulentum* X79714. **11. plant-1** H3.1: [polyA] Plantae, dicot *M. sativa* A38309, M35867, S04520, U09459, X13673, X13674, X13675, X13677; *Pisum sativum,* pea A02631, P02300; monocot *Triticum sativum,* wheat A26014, S00373, X00937. **12. plant-2** H3.1: [polyA] Plantae, dicot *A. thaliana* M17130, M17131, M35387, P05329 {40, 49, 132 typing errors}, P05330 {49, 55 typing errors}, S06250; *Petroselinum crispum,* parsley M77493, M77494, M77495; monocot *Z. mays* M13378, M13379, M35388, M36658, P05203; *Oryza sativa,* rice A25564, M15664 {86, 90, 98 sequencing errors}, P08860, S04099, X13678, X13680. Excluded pseudogene: *O. sativa* X13679. *Encephalartos altensteinii* A23604, and *Encephalartos caffer* C61286 cycad histone protein H3 forms are incompletely defined and were excluded. **13. S. pombe-1** H3: [polyA] Fungi, Hemiascomycetes *Schizosaccharomyces pombe,* fission yeast E27399, P09988, X05222, X05223. **14. S. pombe-2** H3: [polyA] Fungi, Hemiascomycetes *S. pombe* P10651, X05224. **15. S. cerevisiae** H3: [polyA] Fungi, Hemiascomycetes *Saccharomyces cerevisiae* budding yeast A02635, K00900 (excluded: partial sequence with errors), P02303, S45265, X00724, X00725, Z35879; *Kluyveromyces lactis* S07892, X14230. **16. E. nidulans** H3: [polyA] Fungi, Ascomycotina *Emericelli nidulans* [I = 8(2), 44(2)] P23753, S11938, U12630 (corrected for incorrect intron splices), X55548. **17. N. crassa** H3: [polyA] Fungi, Ascomycotina *Neurospora crassa* [I = 71(1)] P07041, S07350, X01612. **18. Ciliata** H3.1 and H3.2: [polyA] Protozoa, Ciliophora *Tetrahymena thermophila* M11142, M11143, M87304, M87504, S42521, S42522; *Glaucoma chattoni* M73209. Partial sequences of *Tetrahymena* species: *T. borealis* P17319, S10263, X17128 {3, 14, typing errors}; *T. pyriformis* X17141; 18 different *Tetrahymena* species P17705, X17125–X17127, X17130–X17140, X17142–X17145. Excluded (partial, divergent sequences): *Tetrahymena* species M73207, M73208, M73210–M73214. **19. Ciliata-R** Replacement H3.3: [polyA] Protozoa, Ciliophora *T. thermophila* M87305, S41501. *T. pyriformis* protein analyses A28852, P15511, and B28852, P15512 conform to complete nucleic acid clones in 20 and 21 {exceptions: 80, 124 and 125}. **20. Plasmodium-1** H3: [polyA] Protozoa, Microspora *Plasmodium falciparum* U14735. **21. Plasmodium-2** H3: [polyA] Protozoa, Microspora *P. falciparum* U15994. **22. Entamoeba** H3: [polyA] Protozoa, Rhizopoda *Entamoeba histolytica* L02418, Q06196 (incomplete sequence alignment). **23. Leishmania** H3: [polyA] Protozoa, Kinetoplastida *Leishmania infantum* S42034, X77591 (incomplete sequence alignment). **24. Trypanosoma** H3: [polyA] Protozoa, Kinetoplastida *Trypanosoma cruzi* L27659, L27660 (incomplete sequence alignment).

Histone H4 Sequences. **25. animal** H4: [Ω-U7] Animalia, Mammalia *H. sapiens* A02636, A93490, D40335, M16707, M60749, P02304, S17086, X00038, X60481, X60482, X60483, X60484, X60486, X60487, X67081, X83548; *B. taurus* A02637, A92050; *Sus scrofa,* pig A02638; *M. musculus* S03426, S03427, V00753, X13235, X13236; *R. norvegicus* A02639, A30772, M24020, M27433, M28409, S02762, X13554; Aves *G. gallus* A02640, B29179, B61321, J00866,

JH0507, M74533, M74534, X02218; *C. moschata* X14732; Pisces *S. gairdneri* A24552, X02916; *Talapia nilotica* X54078; Amphibia *X. laevis* A02641, X00224, X03017, X03018; *Xenopus borealis* J00985; *N. viridescens* J00954; Echiura *U. caupo* S21851, X58895; Mollusca *Sepia officinalis,* cuttle fish A61321; Annelida *P. dumerilii* S11312, X53330. Excluded pseudogene: *H. sapiens* X60485. **26. Echinodermata** H4: [Ω-U7] Animalia, Echinodermata, Echinizoa *P. miliaris* A02642, M10556, X15634; *S. purpuratus* A02643, M32426, V01355, X03952, X06639, X06641; *L. pictus* J01175, X00593; *Holothuria tubulosa* S49485, Z46226; Echinodermata, Asterozoa *P. brevispinus* S20666, X54112; *P. ochraceus* S20670, X54113; *P. helianthiodes* S20668, X54114; *S. stimpsoni* P27996, S20677, X54115 {83 sequencing error}; Nematoda *C. elegans* JS0314, P02306, S04240, X15634. **27. Arthropoda** H4: [Ω-U7] Animalia, Insecta *D. melanogaster* A02644, P02307, S10098, X14215; *D. hydei* S09656, S21937, X17072, X52576; *D. virilis* M32469; *C. thummi* S18003, S40437, X56335, X72803; Crustacea *T. californicus* M84797, X52393. **28. Tunicata** H4: [Ω-U7] Animalia, Tunicata *S. plicata* JN0688, S64499 {34 codon translation error}. **29. Cnidaria** H4: [Ω-U7] Animalia, Cnidaria *A. formosa* L11067, P53059, S67324 {if position 21 is sequencing error, this is sequence 25}. **30. Pyrenomonas** H4: [Ω-polyA] Chromophycota, Cryptophyceae *Pyrenomonas salina* X77806. **31. Physarum** H4: [Ω-u?] Myxomycota *Physarum polycephalum* [I = 45(2)] A27859, P04915, S10075, S10076, X00449 {splice location error}, X15141, X15142, Y00366 {48 codon translation error}. **32. Volvox** H4: [Ω-u?] Chlorophycota *V. carteri* P08436, S00939, X06963, X06964. **33. Chlamydomonas** H4: [Ω-u?] Chlorophycota *C. reinhardtii* L41841, U16724, U16725, U16825. **34. plant-1** H4: [polyA] Plantae, dicot *A. thaliana* M17132, M17133, S06904, Z17567; *M. sativa* P02308; *P. sativum* A02646, U10042; monocot *T. aestivum* A02645, A24967, M12277 {4-sequencing error}, P02308, X00043; *Z. mays* A25642, M13370, M13377, M36659, T14777 (EST), T14749 (EST); *O. sativa* D10397; *L. temulentum* X79715. **35. plant-2** H4: [polyA] Plantae, dicot *L. esculentum* P35057, S32769, S32770, X69179, X69180. **36. P. chrysosporium** H4: [polyA] Fungi, Basidiomycotina *Phanerochaete chrysosporium* P35058, S25638, S25639, Z15134 [I = 9(1), 28(2), 65(1)], Z15135 [I = 9(1), 33(1), 78(3)]. **37. S. cerevisiae** H4: [polyA] Fungi, Hemiascomycetes *S. cerevisiae* A02647, K03154, P02309 {45-protein-sequence lag error}, S07914, X00724, X00725, Z35878. **38. S. pombe** H4: [polyA] Fungi, Hemiascomycetes, *S. pombe* D27399, P09322, X05222, X05223, X05224. **39. E. nidulans-1** H4: [polyA] Fungi, Ascomycotina *E. nidulans* [I = 3(2)] P23751, S11940, U12631, X55550. **40. E. nidulans-2** H4: [polyA] Fungi, Ascomycotina *E. nidulans* [I = 3(2), 95(2)] P23750, S11939, U12630, X55549. **41. N. crassa** H4: [polyA] Fungi, Ascomycotina *N. crassa* [I = 3(2), 49(2)] P04914, S07913, X01611. **42. Ciliata** H4: [polyA] Protozoa, Ciliophora *Oxytricha nova* JS0154, M24411, P18836; *Stylonychia lemnae* S14184, S14185, X16018, X16019. **43. Tetrahymena-1** H4: [polyA] Protozoa, Ciliophora *T. thermophila* A02650, A21029, A25875, X00417, X04755; *T. pyriformis* A02649, P02311; 15 different *Tetrahymena* species (partial sequences) M73208, M73211, M73212, M73213, M73214, X17127, X17128, X17129, X17131, X17132, X17134, X17136, X17141, X17144, X17145. **44. Tetrahymena-2** H4: [−] Protozoa, Ciliophora *T. pyriformis* A02648, P02310. Excluded (partial, divergent sequences): *Tetrahymena* species M73207, M73210, X17125, X17126, X17130, X17133, X17135, X17137–X17140, X17142, X17143; *G. chattoni* M73209. **45. Entamoeba** H4: [polyA] Protozoa, Rhizopoda *E. histolytica* L35898, P40287, S52264, X84009, X84010 (incomplete sequence alignment).

Results

Phylogenetic Analysis of Histone H3 and H4 Proteins

Sequences for histone H3 (Fig. 1) and H4 (Fig. 2) were compiled with corrections for database entry errors (see

solidissima, clam M17876. **5. Echinodermata-R** Replacement H3.3: [polyA] Animalia, Echinodermata *P. lividus* X76081. **6. Nematoda-1** major H3: [–] Animalia, Nematoda *Caenorhabditis elegans* P08898 (major protein variant). **7. Nematoda-2** minor H3: [Ω-U7] Animalia, Nematoda *C. elegans* X15634. Entries 6 and 7 do not fully represent histone H3 complexity in *C. elegans*: Z38112 (EST), (Vanfleteren et al. 1989). **8. Cnidaria** H3: [Ω-U7] Animalia, Cnidaria *Acropora formosa,* coral JQ0757, L11067, M60509, P22943, S67324. **9. algae** H3: [Ω-u?] Chlorophycota *Volvox carteri* S00940, X06963 [I = 46(2)], X06964 [I = 46(1)]; *Chlamydomonas reinhardtii* L41841 {90, 113 sequencing errors}, U16724, U16725, U16825. **10. plant-R** Replacement H3-III, H3.2 and H3.3: [polyA] Plantae, dicot *Arabidopsis thaliana* [I = 5′UTR, 24(2), 49(3), 80(1) (gene 1); 5′UTR, 24(2), 49(3) (gene 2)] S24346, X60429; *Medicago sativa,* alfalfa [I = 24(1), 49(3), 80(1)] B38309, S04521, U09458, U09460, U09461, U09462, U09463, U09464, U09465, X13676; *Lycopersicon esculentum,* tomato X83422; monocot *Hordeum vulgare,* barley A02632, P06353; *Zea mays,* corn T14800 (EST), T23317 (EST), T23357 (EST); *Lolium temulentum* X79714. **11. plant-1** H3.1: [polyA] Plantae, dicot *M. sativa* A38309, M35867, S04520, U09459, X13673, X13674, X13675, X13677; *Pisum sativum,* pea A02631, P02300; monocot *Triticum sativum,* wheat A26014, S00373, X00937. **12. plant-2** H3.1: [polyA] Plantae, dicot *A. thaliana* M17130, M17131, M35387, P05329 {40, 49, 132 typing errors}, P05330 {49, 55 typing errors}, S06250; *Petroselinum crispum,* parsley M77493, M77494, M77495; monocot *Z. mays* M13378, M13379, M35388, M36658, P05203; *Oryza sativa,* rice A25564, M15664 {86, 90, 98 sequencing errors}, P08860, S04099, X13678, X13680. Excluded pseudogene: *O. sativa* X13679. *Encephalartos altensteinii* A23604, and *Encephalartos caffer* C61286 cycad histone protein H3 forms are incompletely defined and were excluded. **13. S. pombe-1** H3: [polyA] Fungi, Hemiascomycetes *Schizosaccharomyces pombe,* fission yeast E27399, P09988, X05222, X05223. **14. S. pombe-2** H3: [polyA] Fungi, Hemiascomycetes *S. pombe* P10651, X05224. **15. S. cerevisiae** H3: [polyA] Fungi, Hemiascomycetes *Saccharomyces cerevisiae* budding yeast A02635, K00900 (excluded: partial sequence with errors), P02303, S45265, X00724, X00725, Z35879; *Kluyveromyces lactis* S07892, X14230. **16. E. nidulans** H3: [polyA] Fungi, Ascomycotina *Emericelli nidulans* [I = 8(2), 44(2)] P23753, S11938, U12630 (corrected for incorrect intron splices), X55548. **17. N. crassa** H3: [polyA] Fungi, Ascomycotina *Neurospora crassa* [I = 71(1)] P07041, S07350, X01612. **18. Ciliata** H3.1 and H3.2: [polyA] Protozoa, Ciliophora *Tetrahymena thermophila* M11142, M11143, M87304, M87504, S42521, S42522; *Glaucoma chattoni* M73209. Partial sequences of *Tetrahymena* species: *T. borealis* P17319, S10263, X17128 {3, 14, typing errors}; *T. pyriformis* X17141; 18 different *Tetrahymena* species P17705, X17125–X17127, X17130–X17140, X17142–X17145. Excluded (partial, divergent sequences): *Tetrahymena* species M73207, M73208, M73210–M73214. **19. Ciliata-R** Replacement H3.3: [polyA] Protozoa, Ciliophora *T. thermophila* M87305, S41501. *T. pyriformis* protein analyses A28852, P15511, and B28852, P15512 conform to complete nucleic acid clones in 20 and 21 {exceptions: 80, 124 and 125}. **20. Plasmodium-1** H3: [polyA] Protozoa, Microspora *Plasmodium falciparum* U14735. **21. Plasmodium-2** H3: [polyA] Protozoa, Microspora *P. falciparum* U15994. **22. Entamoeba** H3: [polyA] Protozoa, Rhizopoda *Entamoeba histolytica* L02418, Q06196 (incomplete sequence alignment). **23. Leishmania** H3: [polyA] Protozoa, Kinetoplastida *Leishmania infantum* S42034, X77591 (incomplete sequence alignment). **24. Trypanosoma** H3: [polyA] Protozoa, Kinetoplastida *Trypanosoma cruzi* L27659, L27660 (incomplete sequence alignment).

Histone H4 Sequences. **25. animal** H4: [Ω-U7] Animalia, Mammalia *H. sapiens* A02636, A93490, D40335, M16707, M60749, P02304, S17086, X00038, X60481, X60482, X60483, X60484, X60486, X60487, X67081, X83548; *B. taurus* A02637, A92050; *Sus scrofa,* pig A02638; *M. musculus* S03426, S03427, V00753, X13235, X13236; *R. norvegicus* A02639, A30772, M24020, M27433, M28409, S02762, X13554; Aves *G. gallus* A02640, B29179, B61321, J00866, JH0507, M74533, M74534, X02218; *C. moschata* X14732; Pisces *S. gairdneri* A24552, X02916; *Talapia nilotica* X54078; Amphibia *X. laevis* A02641, X00224, X03017, X03018; *Xenopus borealis* J00985; *N. viridescens* J00954; Echiura *U. caupo* S21851, X58895; Mollusca *Sepia officinalis,* cuttle fish A61321; Annelida *P. dumerilii* S11312, X53330. Excluded pseudogene: *H. sapiens* X60485. **26. Echinodermata** H4: [Ω-U7] Animalia, Echinodermata, Echinizoa *P. miliaris* A02642, M10556, X15634; *S. purpuratus* A02643, M32426, V01355, X03952, X06639, X06641; *L. pictus* J01175, X00593; *Holothuria tubulosa* S49485, Z46226; Echinodermata, Asterozoa *P. brevispinus* S20666, X54112; *P. ochraceus* S20670, X54113; *P. helianthiodes* S20668, X54114; *S. stimpsoni* P27996, S20677, X54115 {83 sequencing error}; Nematoda *C. elegans* JS0314, P02306, S04240, X15634. **27. Arthropoda** H4: [Ω-U7] Animalia, Insecta *D. melanogaster* A02644, P02307, S10098, X14215; *D. hydei* S09656, S21937, X17072, X52576; *D. virilis* M32469; *C. thummi* S18003, S40437, X56335, X72803; Crustacea *T. californicus* M84797, X52393. **28. Tunicata** H4: [Ω-U7] Animalia, Tunicata *S. plicata* JN0688, S64499 {34 codon translation error}. **29. Cnidaria** H4: [Ω-U7] Animalia, Cnidaria *A. formosa* L11067, P53059, S67324 {if position 21 is sequencing error, this is sequence 25}. **30. Pyrenomonas** H4: [Ω-polyA] Chromophycota, Cryptophyceae *Pyrenomonas salina* X77806. **31. Physarum** H4: [Ω-u?] Myxomycota *Physarum polycephalum* [I = 45(2)] A27859, P04915, S10075, S10076, X00449 {splice location error}, X15141, X15142, Y00366 {48 codon translation error}. **32. Volvox** H4: [Ω-u?] Chlorophycota *V. carteri* P08436, S00939, X06963, X06964. **33. Chlamydomonas** H4: [Ω-u?] Chlorophycota *C. reinhardtii* L41841, U16724, U16725, U16825. **34. plant-1** H4: [polyA] Plantae, dicot *A. thaliana* M17132, M17133, S06904, Z17567; *M. sativa* P02308; *P. sativum* A02646, U10042; monocot *T. aestivum* A02645, A24967, M12277 {4-sequencing error}, P02308, X00043; *Z. mays* A25642, M13370, M13377, M36659, T14777 (EST), T14749 (EST); *O. sativa* D10397; *L. temulentum* X79715. **35. plant-2** H4: [polyA] Plantae, dicot *L. esculentum* P35057, S32769, S32770, X69179, X69180. **36. P. chrysosporium** H4: [polyA] Fungi, Basidiomycotina *Phanerochaete chrysosporium* P35058, S25638, S25639, Z15134 [I = 9(1), 28(2), 65(1)], Z15135 [I = 9(1), 33(1), 78(3)]. **37. S. cerevisiae** H4: [polyA] Fungi, Hemiascomycetes *S. cerevisiae* A02647, K03154, P02309 {45-protein-sequence lag error}, S07914, X00724, X00725, Z35878. **38. S. pombe** H4: [polyA] Fungi, Hemiascomycetes, *S. pombe* D27399, P09322, X05222, X05223, X05224. **39. E. nidulans-1** H4: [polyA] Fungi, Ascomycotina *E. nidulans* [I = 3(2)] P23751, S11940, U12631, X55550. **40. E. nidulans-2** H4: [polyA] Fungi, Ascomycotina *E. nidulans* [I = 3(2), 95(2)] P23750, S11939, U12630, X55549. **41. N. crassa** H4: [polyA] Fungi, Ascomycotina *N. crassa* [I = 3(2), 49(2)] P04914, S07913, X01611. **42. Ciliata** H4: [polyA] Protozoa, Ciliophora *Oxytricha nova* JS0154, M24411, P18836; *Stylonychia lemnae* S14184, S14185, X16018, X16019. **43. Tetrahymena-1** H4: [polyA] Protozoa, Ciliophora *T. thermophila* A02650, A21029, A25875, X00417, X04755; *T. pyriformis* A02649, P02311; 15 different *Tetrahymena* species (partial sequences) M73208, M73211, M73212, M73213, M73214, X17127, X17128, X17129, X17131, X17132, X17134, X17136, X17141, X17144, X17145. **44. Tetrahymena-2** H4: [–] Protozoa, Ciliophora *T. pyriformis* A02648, P02310. Excluded (partial, divergent sequences): *Tetrahymena* species M73207, M73210, X17125, X17126, X17130, X17133, X17135, X17137–X17140, X17142, X17143; *G. chattoni* M73209. **45. Entamoeba** H4: [polyA] Protozoa, Rhizopoda *E. histolytica* L35898, P40287, S52264, X84009, X84010 (incomplete sequence alignment).

Results

Phylogenetic Analysis of Histone H3 and H4 Proteins

Sequences for histone H3 (Fig. 1) and H4 (Fig. 2) were compiled with corrections for database entry errors (see

```
                                        N-helix           helix-1              helix-2                      helix-3
                                        /\/\/ \/\         /\/\ /\/\/            \/\ /\/\/\/\ /\/\/\/\ /\        \/\/\/\/ /\
            1        10        20        30        40        50        60        70        80        90       100       110       120       130       3'
 1 mammal   ARTKQTARKS TGGKAPRKQL ATKAARKSAP ATGGVKKPHR YRPGTVALRE IRRYQKSTEL LIRKLPFQRL VREIAQDFKT DLRFQSSAVM ALQEACEAYL VGLFEDTNLC AIHAKRVTIM PKDIQLARRI RGERA   Ω-U7
 2 animal   .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .....   Ω-U7
 3 Echinodermata ......  .......... .......... .......... .......... .......... .......... ....S..... E......... ...S...... .......... .......... .......... .....   Ω-U7
 4 animal-R .......... .......... .......... ......S... .......... .......... .......... .......A.IG ...S...... .......... .......... .......... .......... .....   polyA
 5 Echinodermata-R .......... .......... ......S... .......... .......... .......... E...A.IG ...S...... .......... .......... .......... .......... .....   polyA
 6 Nematoda-1 .......... .......... ......S... .......... .......... ....A..... .......... .......... ....A..... .......... .......... .......... .......... .....   -
 7 Nematoda-2 .......... .......... ......S... .......... .......... ....RA.... .......... .......... ....A..... .......... .......... .......... .......... .....   Ω-U7
 8 Cnidaria .......... .......A.. .......... .......... .......... .......... .......... .......... ...S...... .......... .......... .......... .......... .....   Ω-U7
 9 algae    .......... .......... ..T....... .......... .......... ......K... .......... .......... Q..L...... .......... .......... .......... .......... .....   Ω-u?
10 plant-R  .......... .......... ....T..... .......... .......... ......K... .......... .......... H..L...... .A........ .......... .......... .......... .....   polyA
11 plant-1  .......... .......... .......... .......F... .......... ......K... .......... .......... ...S...... .A........ .......... .......... .......... .....   polyA
12 plant-2  .......... .......... .......... .......F... .......... ......K... .......... .......... .......A.. .A........ .......... .......... .......... .....   polyA

13 S.pombe-1 .......... .......... .S....A... .......... .......... ......K... .......... ..IG ...... V.. ....... .S.... ...G...Q ..M....L ...S polyA
14 S.pombe-2 .......... .......... .S....A... .......... ...P...... ......K... .......... ..IG ...... V.. ....... .S.... ...G...Q ..M....L ...S polyA
15 S.cerevisiae .......... .......... .S..A...S. .......K... .......F.. .......... ..IG ...SV... .S.....A .. .....Q K..K....L ...S polyA
16 E.nidulans .......... .......... .S..A...S. .......K... .......... .......... ....S..... ..IG ...SV... .S........ .....Q S....... ...L ...S polyA
17 N.crassa  .......... .......... .S..A...S. .......K... .......... .......... ....S..... ..IG L...SV.S.. .S........ .....Q S....... ...L ...N polyA

18 Ciliata   .......... ..A....... .S........ ...vs...I.. ...F...... ......K....D. ........ ..D..HE..A E ......L ...A..... .......... ....R..... T..M..... ....F polyA
19 Ciliata-R .......... ..v....... .......... ...vS...K F ...F..... ......K...T.D. ........ ..D..MEM.S .I...Q.IL ...A..... .......... ....R..... T..LH..... ....F polyA
20 Plasmodium-1 E.....E... .......... .S.EE..... VST.I..... .......... .....KS...D. ........ ...EY..... ...Q.L... ...A..... .......... .......I I .......... .... polyA
21 Plasmodium-2 .......... .......... .......... VST.I..... .......... .......KF..D. ........ ...EY..... ...Q.L... ...A..... .......... .......... ..M...... .... polyA
22 Entamoeba .... ghierpsnksakavknv .F..K. mlskdst .. krah .A..T. .KVL.R... .L..A..A. ........ ..VS.S .A.IS .....A....... .........I .......M..... ....T polyA
23 Leishmania srtk etarakrtit skkskkapsg asgvkrshrr w...C.I... .KF...S. .QCA..... .... GVERQ.E G........I. .....T...I .S.MA....A C......... Q .......L.L ....H polyA
24 Trypanosoma srsket arskrtitsk kskkpprplv rppreartpa v......... ..QF.R..D. .LQ.A..... ...VSGAQ.E G.......IL .A..T.S.V .S.LA...RA C..SG....Q ....H..LCL ..... polyA
```

Fig. 1. Histone H3 protein sequences, in *single-letter amino acid code*, compiled from GenBank and PIR databases as described in Materials and Methods with detailed information on the descriptive sequence name, combined sequence identities and species, and applied corrections. Mature protein sequences are aligned with sequence 1, except for the divergent parts of sequences 22–24, which are shown in *lowercase*. Sequence 1 is *numbered* as in the mature protein. α-Helical domains N-helix, helix-1, helix-2, and helix-3 are marked (Arents and Moudrianakis 1995). A *dot (.)* indicates identity to sequence 1; *3'* indicates features identified in the 3'UTR of the gene transcript: Ω indicates a stem-loop secondary RNA structure; *U7* indicates a downstream processing sequence that interacts with U7 snRNA; *u?* indicates the suggested presence of a downstream processing sequence without the U7 snRNA-related sequence; *polyA* marks the demonstrated presence of mRNA polyadenylation; − marks the absence of 3'UTR information.

```
                                     N-helix     helix-1               helix-2                  helix-3
                                     /\          \/\/\/\/               \/\ /\/\/\/\ /\/\/\/\ /\/\    /\/\/\/ /\/\
            1        10        20        30        40        50        60        70        80        90       100       3'
25 animal   SGRGKGGKGL GKGGAKRH RK VLRDNIQGIT KPAIRRLARR GGVKRISGLI YEETRGVLKV FLENVIRDAV TYTEHAKRKT VTAMDVVYAL KRQGRTLYGF GG   Ω-U7
26 Echinodermata ...... .......... .......... .......... .......... .......... .......... .....C.... .......... .......... ..   Ω-U7
27 Arthropoda T......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..   Ω-U7
28 Tunicata  .......... .......... .......... .......... .......... .......... .......... ....L..... .......... .......... ..   Ω-U7
29 Cnidaria  .......... ...... ..I .......... .......... .......... .......... .......... .......... .......... .......... ..   Ω-U7
30 Pyrenomonas ....... .......... .......... .......... .......... .....S.... .......... .....R.... .......... .......... ..   Ω-polyA
31 Physarum  .......... .......... ....KT.... .......... .......T.. .......... .......... .....R.... .......... .......... ..   Ω-u?
32 Volvox    .......... .......... .......... .......... ....T...N. .......... .......... .....R.... .......... .......... ..   Ω-u?
33 Chlamydomonas ...... .......... .......... .......... ....T..T.. .......S.. .......... .....R.... .......... .......... ..   Ω-u?
34 plant-1   .......... .......... .......... .......... .......I.. .......S.. .......... .....R.... .......... .......... ..   polyA
35 plant-2   .......... .......... .......... .......... .......I.. .......S.. .......... .....R.... .......... .......... ..   polyA

36 P.chrysosporium ...... ....I..... .......... .......... .......I.. .......S.. ...L...... .....S.... .......... ..S...... .A   polyA
37 S.cerevisiae ...... ....I..... .......... ...V.A..S .S...S.... .SL...... .....S.... .......... .......... ..   polyA
38 S.pombe   .......... ....I..... .......... .......... .......... .SL...... .....S.... .......... .......... ..   polyA
39 E.nidulans-1 ..... ....I..... .......... .A.V ...A..L .......... .SL...S.. .....S.... .......I.. .......... ..   polyA
40 E.nidulans-2 ..... ....I..... .......... .AM. ...S ...S .......T ...G.... .SL...... .....R.... .......... .......... ..   polyA
41 N.crassa  T......... ....I..... .......... .AM. ...T ...G.... .SL...... .....R.... .......... .......... ..   polyA

42 Ciliata   A....V...Y ..V.....TK. S.KET.M... .......... .......S. ...N..RS .......S. .......... ...L...... .......... ..   polyA
43 Tetrahymena-1 A... ...M ..V...S. SNKAS.E... .......... .......SF. .DDS.Q...S .....V.... .....R.... .......... .......... ..   polyA
44 Tetrahymena-2 A... ......M ..V...SKR SNKAS.E... .......... .......SF. .DDS.Q...S .......... .....R.... .......... .......... ..   -
45 Entamoeba atdtg .....^...krirtk tqqdalk ... .......... ......N.AV .D...N...Q ..Q....S. .......R.... .........Y S   polyA
                    vtlgkgskgakas
```

Fig. 2. Histone H4 protein sequences, compiled and marked as described for Fig. 1, with alignment on sequence 25.

Materials and Methods). The sequences of animals and plants, of fungi, and of protista formed groups with 90%, 84%, and 60% residue conservation, alike for H3 and H4 histones (Table 1), reflecting their intimate interaction within the nucleosome tetramer (Arents et al. 1991; Arents and Moudrianakis 1995).

Neighbor-joining analysis of histone H3 sequences produces a phylogenetic tree with mostly low-confidence branches (Fig. 3) (Kumar et al. 1993). This is caused by the limited number of distinct histone H3 sequences (Hillis et al. 1994). The tree structure is similar to that described before (Thatcher et al. 1994) but is more consistent with generally accepted divergence of phyla due to the exclusion of entries with database errors. The long branch lengths for the protist species are consistent with a higher rate of sequence divergence for these organisms,

which is reflected in their more divergent chromatin structures (Sadler and Brunk 1992; Fodinger et al. 1993; Michalon et al. 1993; Schlimme et al. 1993; Lanzer et al. 1994; Soto et al. 1994). This leads to branch clustering which does not necessarily reflect the true evolutionary relationship. The instability of the protist data was also reflected in the changing branch point for the group of protist sequences when raw differences or distinct algorithms for distance calculation were used for all or subsets of sequences (Kumar et al. 1993). However, the branching pattern for the animal, plant, and fungal H3 sequences was stable, remarkable in view of the low bootstrap confidence probabilities of many branches (Fig. 3). The result suggests that animal replacement H3.3 genes (sequences 4–5) diverged from the replication forms (sequences 1–3, 6–8) prior to the time that

Table 1. Protein sequence conservation of histones H3 and H4

Invariant residues	Histone H3 (135 residues)			Histone H4 (102 residues)		
	Sequences	Residues	%	Sequences	Residues	%
Animals, plants	1–12	122	90	25–35	92	90
Fungi	13–17	113	84	36–41	86	84
Protista[a]	18–19, 22	77	57	42–45	64	63

[a] Comparison is limited to protists for which histone H3 and H4 sequences are known.
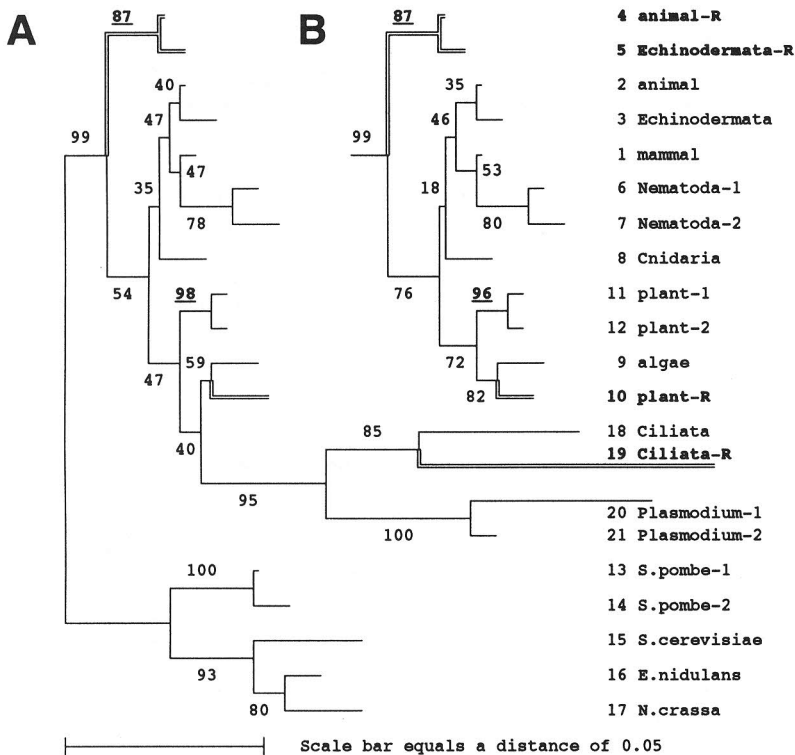
Fig. 3. Phylogenetic tree for histone H3 sequences, generated by 1,000 bootstrap neighbor-joining analyses of sequences 1–21 by the program MEGA as described in Materials and Methods with bootstrap confidence levels given along horizontal branches (**A**). Each branch is identified by the sequence label used in Fig. 1. The tree is unrooted. Inset **B** shows the tree for sequences 1–12 in an analysis of sequences 1–17. The result for sequences 13–17 was identical to that obtained in **A**. The total evolutionary divergence between two sequences is represented by the sum of the horizontal branch lengths between the two sequences. The *scale bar* for a *p*-distance of 0.05 is shown. Vertical distances are for illustration purposes only. Replacement histone H3 branches that result from H3 gene duplication and divergence rather than from histone H3 sequence divergence during speciation are marked by *double lines* and the branch labels are shown in *bold*. The branches with *underlined* bootstrap values are discussed.

primitive multicellular metazoa like Cnidaria (coral) developed.

A replacement histone H3 variant is present in all higher, multicellular plants analyzed (Waterborg 1991, 1992). It is absent in the unicellular green alga *Chlamydomonas* (Waterborg et al. 1995). The histone H3 complexity in primitive plants has not been analyzed. The neighbor-joining analysis (Fig. 3) suggests that a plant replacement histone H3 form arose during or after divergence of protophyta into plants and green algae (bootstrap confidence level 96–98%).

High neighbor-joining bootstrap values reflect a degree of confidence that the observed relationship is real. Maximum parsimony analysis does the same (Fig. 4). In addition, it allows one to determine the likelihood that divergent branches might share a common origin. In only 0.8% of all parsimony analyses, a common origin was observed between the animal (sequences 4–5) and plant (sequence 10) replacement H3 histones. This provides strong support for the inference that the replacement his-

tone H3 forms in animals and in plants arose independently by divergent duplication of existing histone H3 genes.

Neighbor-joining analysis of the compiled histone H4 sequences (Fig. 2) produced only a low-confidence tree (Fig. 5). It is consistent with the histone H3 analyses and with generally accepted divergence among phyla.

The Presence of Introns in Histone H3 and H4 Genes

The high degree of sequence conservation in histone proteins reflects the structural and functional constraints imposed by the nucleosome (Arents et al. 1991; Arents and Moudrianakis 1995). Features of the histone genes have been quite variable, in codon usage (Marzluff 1989; Thatcher et al. 1994), partially reflecting major changes in C+G content of whole genomes (Chaboute et al. 1993), in gene copy numbers, and in the overall organization of histone genes as single genes, tight clusters,
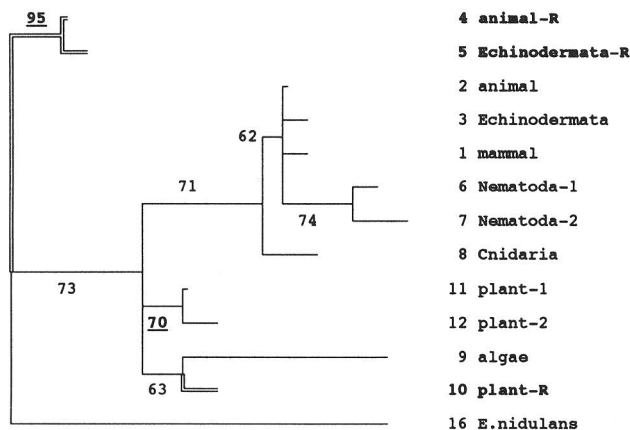
Fig. 4. Parsimony tree of histone H3 sequences 1–12 with sequence 16 as outgroup (Fig. 1) obtained from 250 replicates using the heuristic algorithm of the PAUP program (see Materials and Methods). Bootstrap confidence levels above 50% are given along the horizontal branches. Branches with lower confidence levels were collapsed. The frequency of trees in which sequences 4, 5, and 10 share a common origin was 0.8%. No instance was recorded showing that sequences 5 and 10 or 4 and 10 share a common origin. All other details are as in Fig. 3.

large arrays of genes, or as fully dispersed multiple copies (Old and Woodland 1984; Chaboute et al. 1993). However, the pattern of introns in histone genes and the processing features of histone gene transcripts are highly conserved and provide support for the notion that replacement histone H3 genes originated more than once.

Introns are rare in histone H3 and H4 genes (Table 2) and absent in the cell-cycle-regulated histone genes of all higher eukaryotes. Acquisition of one or more introns into the archetypical intron-free histone gene has been rare (Figs. 6 and 7). Only in four instances has an intron insertion occurred into any cell-cycle-dependent histone H3 gene (Table 2): twice in *E. nidulans,* once in *N. crassa* and once in *V. carteri* (Fig. 6). The limitation to lower eukaryotes, almost exclusively to fungi, is also observed for histone H4 genes. Single intron insertions are inferred for *P. polycephalum, P. chrysosporium,* and for a common ancestor of *E. nidulans* and *N. crassa* (Fig. 7). Additional intron insertions must have occurred in the latter three organisms (Table 2).

This pattern contrasts with the inferred insertion of three introns into the ancestral replacement histone H3 gene for man, chicken, and *Drosophila* and the insertion of four introns within the ancestral replacement histone H3 gene of alfalfa and *Arabidopsis* (Fig. 6). The pattern of intron locations reveals that these events are independent (Table 2), in support of the phylogenetic inference that the creations of the animal and plant replacement H3 genes were two independent events, as concluded for histone H2A variant genes in *Tetrahymena* (VanDaal et al. 1990). It is suggested that each case represents the insertion of multiple introns rather than the sequential insertion of single ones.

The Role of Introns in Replacement H3 Genes

This raises the question of whether the insertion of multiple introns would be an obligatory feature if one would need to convert a member of a tightly cell-cycle-regulated histone gene family to become expressed in a cell-cycle-independent manner. Theoretically one can argue the initial advantage of insertion of multiple introns into a newly duplicated copy of a gene. It would insulate the gene from recombinational homogenization with other members of the gene family and allow divergent differentiation of function, if allowed or required. However, over time, sequence drift could naturally isolate the divergent gene and obviate the need for the introns. The conserved presence of the introns argues that they are required.

Retaining *all* introns apparently has not been required. The loss of the 5'UTR intron from the alfalfa replacement H3 gene was used to identify features that might be essential for a replacement histone H3 gene and that might have been retained in a different form (Robertson 1994; Robertson et al. 1995). The 5'UTR of the alfalfa replacement H3 gene is a variable repeated pattern of cytosine and thymidine nucleotides which was called a polypyrimidine motif (Robertson et al. 1995). Similar polypyrimidine sequences are present in the coding strand of each of the introns of the alfalfa gene (Fig. 8, Ms1) and in the two other known plant replacement H3 genes in *Arabidopsis* (Fig. 8, At1). They are also present in the introns of all known animal replacement H3 genes: two human (Fig. 8, Hs1–Hs2), two chicken (Fig. 8, Gg1–Gg2), and two fruit fly genes (Fig. 8, Dm1, Dh1). In the animal genes, these sequences occur on both DNA strands and thus are seen as polypyrimidine and polypurine motifs. All these genes have an unusually high abundance of clustered polypyrimidine sequences (Table 3). While the cell-cycle-regulated histone H3 sequences are not devoid of general polypyrimidine tracts, their distribution is different (Fig. 8) and they occur at frequencies that are close to theoretical predictions (Table 3).

This comparison suggests that the polypyrimidine motifs of the introns could be required for the appropriate functional expression of replacement histone H3 genes. In a number of cases, the polypyrimidine motifs clearly extend beyond the 5' boundary of the upstream untranslated region of the transcript and include the area of the TATA-containing promoter (Fig. 8). The polypyrimidine sequences in one of the human H3 replacement genes, Hs1 in Fig. 8, have been identified as strong enhancers for transcription initiation (Zhong et al. 1983). We have observed that, on a per-gene basis and adjusted for the differences in cell cycle expression, each replacement H3 gene in alfalfa produces three to five times more transcript (Kapros et al. 1995). Preliminary experimental observations support the notion that these polypyrimidine sequences act as transcriptional enhancers that increase the constitutive expression of genes. Insertion of
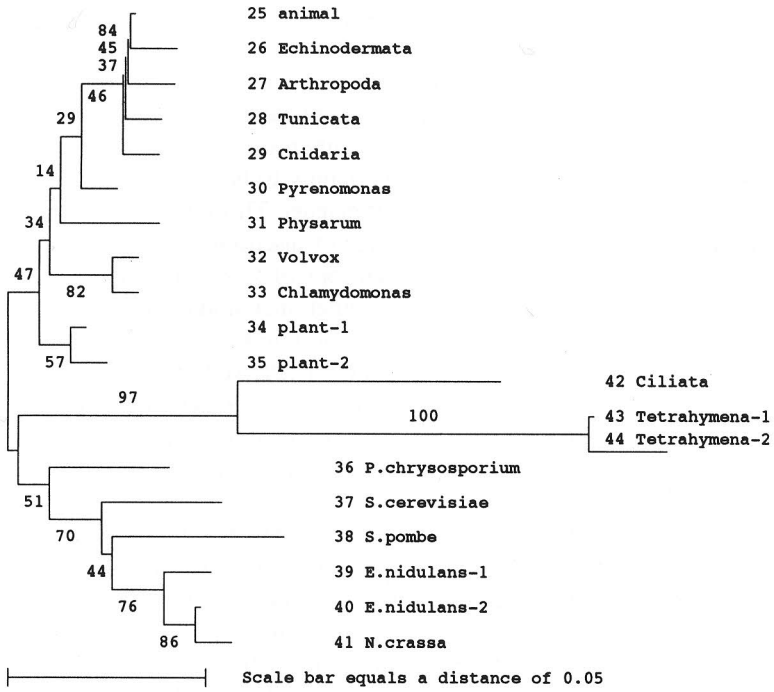
Fig. 5. Phylogenetic tree for histone H4 sequences 25–44 (Fig. 2), generated by 1000 bootstrap neighbor-joining analyses by the program MEGA. All details are as in Fig. 3.

Table 2. Introns in histone H3 and H3 genes

Histone sequence[a]		Species	Clone[b]	Intron 1[c]	Intron 2	Intron 3	Intron 4
H3	**4 animal-R**	*H. sapiens*		5'UTR	42 (3)	94 (1)	
		G. gallus		5'UTR	42 (3)	94 (1)	
		D. melanogaster		5'UTR	—	94 (1)	
		D. hydei		5'UTR	—	94 (1)	
	9 algae	*V. carteri*	X06963	46 (2)			
			X06964	46 (1)			
	10 plant-R	*A. thaliana*	gene 1	5'UTR	24 (2)	49 (3)	80 (1)
			gene 2	5'UTR	24 (2)	49 (3)	—
		M. sativa		—	24 (1)	49 (3)	80 (1)
	16 E. nidulans	*E. nidulans*		8 (2)	44 (2)		
	17 N. crassa	*N. crassa*		71 (1)			
H4	31 Physarum	*P. polycephalum*		45 (2)			
	36 P. chrysosporium	*P. chrysosporium*	Z15134	9 (1)	28 (2)	65 (1)	
			Z15135	9 (1)	33 (1)	78 (3)	
	39 E. nidulans-1	*E. nidulans*	U12631	3 (2)			
	40 E. nidulans-2	*E. nidulans*	U12630	3 (2)	95 (2)		
	41 N. crassa	*N. crassa*		3 (2)	49 (2)		

[a] See Fig. 1 and Fig. 2 for sequences and codon numbering. Replacement histone H3 genes are shown in bold.

[b] See Materials and Methods.

[c] Position of intron is defined by the codon (and codon position) of the first base in the following exon, or is given as inside the 5'UTR; - marks the absence (and inferred loss) of an intron.

the 5'UTR intron of either *Arabidopsis* replacement H3 gene between a strict cell-cycle-regulated histone H4 promoter and a glucuronidase reporter gene in stable transformant cell lines caused a loss of the specificity for expression in cycling meristem cells in *planta* and increased gene expression overall eight- to tenfold (Chaubet, Waterborg, and Gigot, unpublished results). The polypyrimidine sequences of the intronless 5'UTR of the alfalfa replacement H3 gene have a similar effect in stable transformants but they have little effect on transient expression (Dudits, Kapros, and Waterborg, unpub-

lished results). This suggests that the effect of the polypyrimidine sequences may be at the level of facilitated access for transcription factors in chromatin rather than through direct enhancement of transcription.

GAGA factor is a polypyrimidine-binding protein of *Drosophila* that binds to polypyrimidine sequences around promoters in 5'UTRs and introns (Lu et al. 1993; O'Donnell and Wensink 1994), actively displaces nucleosomes (Tsukiyama et al. 1994; Wall et al. 1995), and facilitates transcription by antirepression (Biggin and Tjian 1988; O'Donnell et al. 1994; O'Brien et al.
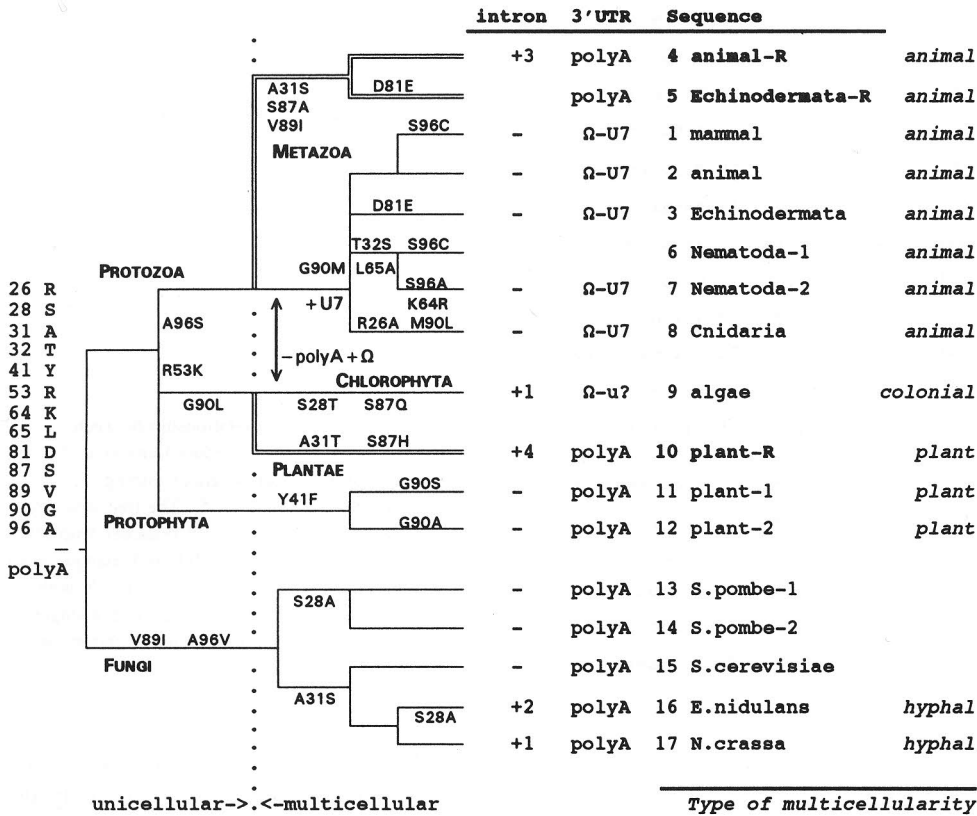
```
                              intron   3'UTR   Sequence

                               +3     polyA    4 animal-R          animal
         A31S      D81E                polyA    5 Echinodermata-R   animal
         S87A
         V89I            S96C          Ω-U7     1 mammal            animal
           METAZOA
                                       Ω-U7     2 animal            animal
                    D81E
                                       Ω-U7     3 Echinodermata     animal
                  T32S  S96C
PROTOZOA      G90M  L65A               -         6 Nematoda-1        animal
                        S96A
            A96S   +U7  K64R           Ω-U7     7 Nematoda-2        animal
                       R26A  M90L
            R53K      ─polyA+Ω         Ω-U7     8 Cnidaria          animal
                       CHLOROPHYTA
            G90L      S28T   S87Q      +1      Ω-u?    9 algae       colonial
                    A31T   S87H
              PLANTAE                  +4     polyA   10 plant-R     plant
              Y41F          G90S       -      polyA   11 plant-1     plant
PROTOPHYTA                  G90A       -      polyA   12 plant-2     plant

                    S28A               -      polyA   13 S.pombe-1
                                       -      polyA   14 S.pombe-2
      V89I   A96V                      -      polyA   15 S.cerevisiae
      FUNGI     A31S     S28A          +2     polyA   16 E.nidulans  hyphal
                                       +1     polyA   17 N.crassa    hyphal
```

26 R
28 S
31 A
32 T
41 Y
53 R
64 K
65 L
81 D
87 S
89 V
90 G
96 A
—
polyA

unicellular->.<-multicellular   Type of multicellularity

Fig. 6. Evolutionary relationship between histone H3 proteins and deduced amino acid changes, inferred from neighbor-joining (Fig. 3) and parsimony (Fig. 4) phylogenetic analyses. The tree structure was obtained by collapsing all branches with bootstrap confidence levels below 50% and by differentiating between orthologous changes of histone H3 proteins in divergent speciation (shown as __|) and divergent H3 gene duplications within an ancestral species (shown as __||__). Branch length is arbitrary. Major phyla are labeled in *bold capitals*. Divergent protist H3 sequences and gene duplications and divergences without effect on the polypeptide product have been excluded. Replacement histone H3 branches are emphasized by *double-lined* branches and *bold* sequence labels as in Fig. 3. Inferred amino acid changes at all residues, variable among sequences 1–12, are shown along the branches in the format: ancestral residue–residue position–new residue. A possible root is shown for the unrooted neighbor-joining phylogenetic tree in order to list the variable residues of a putative ancestral histone H3 protein (shown as ancestral residue with position number near the root). The changes from the inferred ancestral polyadenylation of histone H3 transcripts to mRNA without polyA (*−polyA*), with regulatory 3'UTR stem-loop features (*+Ω*), and with U7 snRNA processing (*+U7*) are marked. The insertion of one or multiple introns, inferred from Table 1, is marked under *intron*. The absence of introns is shown by a *dash* (-). A *vertical, dotted line* marks the evolutionary change from unicellularity to multicellularity in major phyla. The character of the multicellularity is characterized along the right as animal, colonial (algal), plant and hyphal (fungal).

1995; Wall et al. 1995). Polypyrimidine-binding proteins appear abundant in many organisms (Desjardins and Hay 1993; Li et al. 1994). Figure 8 presents a theoretical prediction of the maximal nucleosomal density of histone H3 genes based on the assumption that GAGA-like proteins exist in these organisms with capabilities like GAGA factor. The predicted overall nucleosomal density of plant and animal replacement H3 genes would be less than two-thirds that of the replication-dependent H3 genes (Table 3) with even less chromatin packaging of promoter and 5'UTR intron sequences (Fig. 8). Nucleosome footprinting of the histone H3 genes of alfalfa is in progress.

If introns, as a vehicle for polypyrimidine sequences, are required for the characteristic constitutive expression of replacement H3 genes of animals and plants, why are they absent in the independent, analogous replacement histone H3.3 gene of *Tetrahymena* (Thatcher et al. 1994)? The macronuclear chromatin, specialized to fa- cilitate transcription by its divergent histones (sequences 18, 43, and 44), may not require derepression to allow constitutive gene expression.

Differences in Transcript Processing Among Histone Genes

Polyadenylation of histone mRNA is often considered a hallmark for cell-cycle-independent gene expression (Mannironi et al. 1989; Marzluff 1989; Standart and Dale 1993; Brocard et al. 1994). However, this conclusion appears valid only for animals (Fig. 6) since transcripts of all histone genes are polyadenylated in fungi (Fahrner et al. 1980), plants (Chaboute et al. 1988; Chaubet et al. 1988; Wu et al. 1989), and protista (Bannon et al. 1983; Puerta et al. 1994; Saint-Guily et al. 1994). The phylogenetic analyses of histone H3 (Fig. 6) and H4 genes (Fig. 7) suggest that polyadenylation of histone tran-
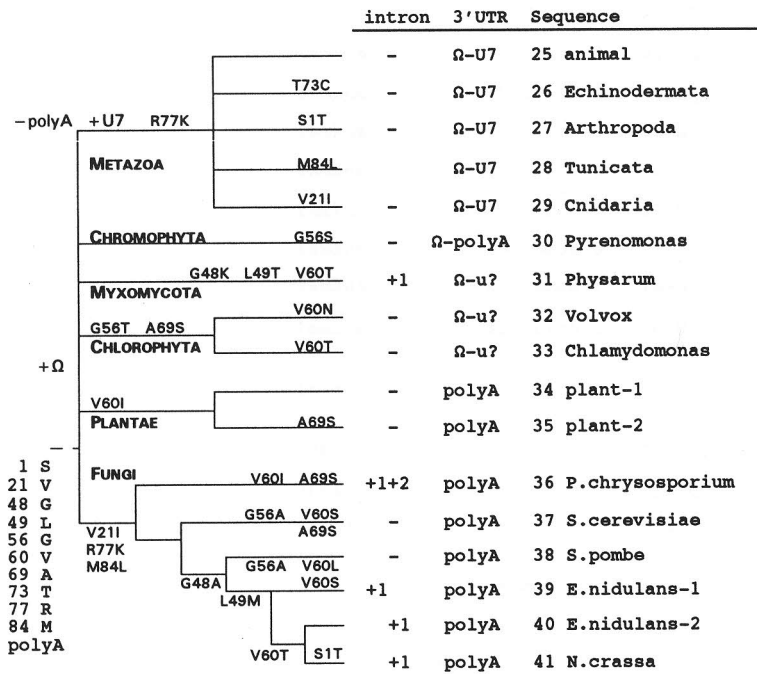
intron	3'UTR	Sequence
–	Ω-U7	25 animal
–	Ω-U7	26 Echinodermata
–	Ω-U7	27 Arthropoda
	Ω-U7	28 Tunicata
–	Ω-U7	29 Cnidaria
–	Ω-polyA	30 Pyrenomonas
+1	Ω-u?	31 Physarum
–	Ω-u?	32 Volvox
–	Ω-u?	33 Chlamydomonas
–	polyA	34 plant-1
–	polyA	35 plant-2
+1+2	polyA	36 P.chrysosporium
–	polyA	37 S.cerevisiae
–	polyA	38 S.pombe
+1	polyA	39 E.nidulans-1
+1	polyA	40 E.nidulans-2
+1	polyA	41 N.crassa

Tree node annotations: –polyA, +U7 R77K, METAZOA, T73C, S1T, M84L, V21I, CHROMOPHYTA G56S, G48K L49T V60T, MYXOMYCOTA, V60N, G56T A69S, CHLOROPHYTA, V60T, +Ω, V60I, PLANTAE, A69S, FUNGI, V60I A69S, G56A V60S, A69S, V21I R77K M84L, G56A V60L V60S, G48A, L49M, V60T S1T

Ancestral positions: 1 S, 21 V, 48 G, 49 L, 56 G, 60 V, 69 A, 73 T, 77 R, 84 M, polyA

Fig. 7. Evolutionary relationship between histone H4 proteins and deduced amino acid changes, inferred from neighbor-joining phylogenetic analysis (Fig. 5). The tree structure was obtained by collapsing all branches with low bootstrap confidence levels. Inferred changes at all positions, variable for sequences 25 to 35, with a putative ancestral histone H4 sequence, changes in transcript processing and insertion of introns are marked as in Fig. 6.

scripts is the ancestral condition. This feature was lost at the start of metazoan divergence (Figs. 6 and 7) and replaced by a U7 snRNP-dependent processing mechanism that yields mature transcripts without polyA tails and with short 3'UTRs. They terminate into a conserved hyphenated stem-loop structure that is part of the U7-dependent recognition and that plays a role in the S phase–dependent processing of transcripts and in mRNA destabilization upon natural or artificial cessation of DNA replication (Heintz 1991; Peltz et al. 1991; Williams et al. 1994). In general, U7-dependent histone processing increases the cell cycle dependency of histone production.

The coexistence of polyA⁺ replacement and polyA⁻ replication H3 genes within all animals for which data are available suggests that the change into nonpolyadenylated transcripts occurred later than the creation of replacement H3 genes. This defines the time of H3 gene duplication and creation of a functional replacement gene creation as prior to species diversification into any animal with nonpolyadenylated histone mRNA, i.e., preceding speciation into corals (Fig. 6).

It remains unclear how old the feature of 3'UTR stem-loop configurations in histone mRNAs might be. It occurs with a high degree of sequence conservation as a hyphenated stem-loop structure in the cell-cycle-dependent histone genes of all animals that contain U7 snRNA (Osley 1991; Marzluff 1992). It is recognizable in green algae but the conserved downstream sequence is different from that associated with U7 action (Fabry et al. 1995). It coexists with polyadenylation in histone transcripts of cryptomonads (Muller et al. 1994) and protists (Soto et al. 1991) and even in replacement linker and H2A histones of animals (Mannironi et al. 1989; Kirsh et

al. 1989). In alfalfa replacement H3 transcripts, a predicted stem-loop structure coincides with the polyadenylation signal sequences (Robertson 1994; Robertson et al. 1995). Their presence or absence in modern histone genes does not correlate with the replacement function. However, the differences in histone transcript termini between plants and green algae support the phylogenetic inference (Fig. 6) that the plant replacement H3 gene originated after plants diverged from green algae.

When Is a Replacement H3 Histone Required?

In a cycling cell, chromatin packaging of DNA is maintained by the periodic supply of new histone protein during S phase. Histones assemble nucleosomes on new DNA and on DNA that has lost nucleosomes due to gene transcription (Björkroth et al. 1988; Waterborg 1993). The development of multicellular organisms with a defined organismal plan inherently involves the creation of noncycling cells. Thus, development of multicellularity, arising from unicellular protist-like ancestral species, requires development of a replication-independent source of histone protein.

Distinct patterns of cell multicellularity have emerged independently. They range from two-dimensional multicellular hyphal arrays of cells in fungi to simple three-dimensional colonies of cells in green algae like *Volvox* and to true three-dimensional organisms that are known as animals and plants. The distinct type of multicellularity in these arose independently and is most easily characterized by presence or absence of a cell wall. Each represents a distinct developmental plan with long-lived,
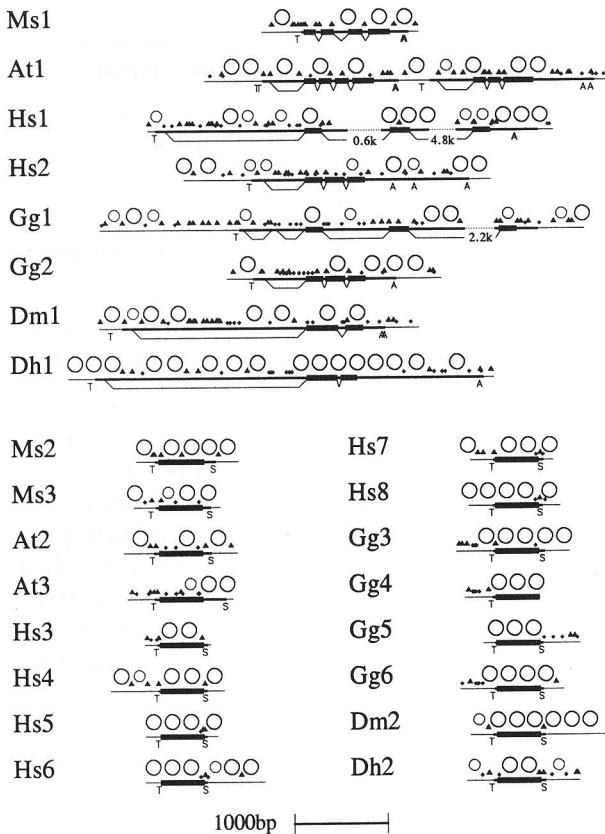
Fig. 8. Comparison of histone H3 gene structures. Shown are all known histone H3 genes from species with genomic clones for replacement histone H3 genes, identified by their GenBank accession numbers: alfalfa *(M. sativa)* Ms1 (U09458), Ms2 (X13673), Ms3 (U09459); *A. thaliana* At1 (X60429), At2 (M17130), At3 (M17131); human *(H. sapiens)* Hs1 (M11454, M77656, X05854–X05857), *Hs2* (Z48950), Hs3 (X00090), Hs4 (M26150), Hs5 (X57128), Hs6 (M60746), Hs7 (X83550), Hs8 (Z46261); chicken *(G. gallus)* Gg1 (M11392, M11667), Gg2 (M11393), Gg3 (Z02218), Gg4 (M61154), Gg5 (M61155), Gg6 (X62291, X62292); *D. melanogaster* Dm1 (X81207), Dm2 (X14215); *D. hydei* Dh1 (X81208), Dh2 (X17072, X52576). Polypyrimidine *(triangles)* and polypurine sequences *(diamonds)* with a minimal length of eight were plotted on the DNA *(thin line)* with transcribed *(medium line)* and translated *(thick line)* sequences, introns and proven or suggested TATA boxes *(T),* polyadenylation signals *(A),* and 3′UTR stem-loop structures *(S).* Sequences were aligned on the ATG start codon. Gaps in sequences are shown as *broken lines,* with estimated length of missing sequence. The *scale* is shown at the bottom. The *top eight lines* represent replacement histone H3 genes. All other lines show known cell-cycle-dependent histone H3 genes. A theoretical nucleosomal packaging of all genes was calculated, assuming that 145 bp of DNA sequence, free of a consecutive polypyrimidine sequence of eight or more nucleotides, will assemble a histone octamer *(small circle),* and that 170 bp, with a repeat of 200 bp, will be packaged by complete nucleosomes *(large circle).*

Table 3. Putative chromatin packaging of histone H3 genes

Genes[a]	Polypyrimidine		Nucleosome
	Length (bp)	Frequency[b]	Frequency[b]
Plant replication H3 Ms2-3 At2–3	9.6 ± 0.8	167 ± 45 (100%)	302 ± 67 (100%)
Plant replacement H3 Ms1 At1	10.2 ± 0.7	131 ± 34 (127%)	463 ± 107 (**65%**)
Animal replication H3 Hs3-8 Gg3-6 Dm2 Dh2	9.6 ± 1.2	230 ± 154 (100%)	266 ± 43 (100%)
Animal replacement H3 Hs1-2 Gg1–2 Dm1 Dh1	10.6 ± 0.8	120 ± 39 (192%)	437 ± 124 (**61%**)
Theoretical	≥8	128	200

[a] Gene codes as used in Fig. 8.
[b] Frequency = number of bp for each element, with relative value between parentheses.

Protein Sequence Variability in Histone H3 and H4

The phylogenetic relationship derived for histone H3 (Fig. 3) and H4 sequences (Fig. 5) was used to analyze the type of amino acid changes that have occurred during speciation and to identify any changes that might be correlated with the replacement function of H3 histones. In both histones, 40% of the changes reflect functional conservation of amino acid size chains. Virtually all other variability is changes in side-chain volume, typified by changes from or to alanine or glycine (Table 4). Based on the structure of the histone octamer (Arents et al. 1991) and the histone clasp structural interactions of the core histones (Arents and Moudrianakis 1995; Baxevanis and Landsman 1996), the few positions with multiple changes are all located on the solvent accessible surface of the nucleosome (Arents and Moudrianakis 1993). The limited variability among the sequences of the higher eukaryotes is fully consistent with maintaining the α-helical domains and turns of the histone clasp as they were defined for the chicken histone octamer (Baxevanis and Landsman 1996). It is likely that the higher variability of protist histones, in the histone clasp region and the amino terminal domains of histones H3 and H4, is the basis for the observed variability in chromatin structure (Sadler and Brunk 1992; Fodinger et al. 1993; Michalon et al. 1993; Schlimme et al. 1993; Lanzer et al. 1994; Soto et al. 1994).

Correlation Between Residues in Histone H3 and Replacement Function

Very few amino acid changes are associated with the emergence of replacement forms of histone H3 in animals and in plants (Fig. 6). Only for two, amino acids 31 and 87, can it be argued that they may play a role in replacement function. Amino acid 31 in histone H3 is always alanine, except in the replacement H3 forms of animals and plants with an deduced change to serine or threonine, respectively (Fig. 6). This represents the gain

permanently noncycling cells. The phylogenetic analysis of histone H3 (Fig. 6) is consistent with the possibility that creation of a replacement H3 form was a necessary element in the development of multicellularity. The implication of this argument is that structural features of the replacement H3 protein may be important and not just its cell-cycle-independent pattern of expression.

Table 4. Variability of amino acids in histones H3 and H4[a]

	Function	Histone H3 (135 residues) Position	Histone H4 (102 residues) Position
Conservation of side-chain function	Hydrophobic	41 54 89 124 130	21 50 84 97
	Hydroxyl	22 32 80	1 73
	Basic	42 53 64	77
	Acidic	81	—
Variation of side-chain size (changes to/from glycine, alanine)		26 28 31 44 65 87 90 91 95 96 98 102 110 114 135	48 56 64 69 83 89 102
Other changes		120 121 125	49 54 60 93

[a] Variable positions in sequences 1–12 (Fig. 1) and 25–35 (Fig. 2) are underlined. Divergent protist sequences 18–24 (Fig. 1) and 42–45 (Fig. 2) are excluded.

of a side-chain function that could participate in hydrogen bonding interactions. Conversely, amino acid 87 changed from serine to a residue without a hydroxyl function, to alanine and histidine in the replacement H3 form of animals and plants, respectively (Fig. 6). In every replacement H3 histone, residues 31 and 87 are completely invariable.

The location of residue 31 within the nucleosome has not yet been determined because it resides in a part of the histone H3 molecule that was not resolved in the histone octamer crystal (Arents et al. 1991). Indirect evidence of proteolytic digestion of nucleosomes and nuclear magnetic resonance (NMR) mobility studies (Matthews and Waterborg 1985) suggest that is located at the surface of the nucleosome. Residue 87 sits at the surface of the histone octamer (Arents et al. 1991), a potential target for interaction with histones, chromatin proteins, or DNA, within the folded structure of chromatin or during transient structural changes that accompany nucleosome assembly (Arents and Moudrianakis 1995; Baxevanis and Landsman 1996).

Preliminary analysis of the two histone H3 forms of alfalfa (Fig. 1, #10 and #11) suggests that differences at positions 31 and 87 may affect what histone H3 can do. These two histones differ only at residues 31, 41, 87, and 90 (Waterborg 1990), and the phylogenetic analysis presented here shows that the changes at residues 41 and 90 occurred after functional histone H3 gene divergence (Fig. 6). Phosphate starvation of alfalfa cells leads to a cell cycle arrest outside of S phase with reduced but significant synthesis of replication H3.1 protein. Analysis of the turnover of this variant protein and of replacement H3.2 protein, synthesized at the same time in these noncycling cells (Waterborg 1993), revealed a more than tenfold difference in protein stability (Waterborg, unpublished results). This reflects the preferential incorporation of the replacement H3.2 protein form into replacement nucleosomes within transcriptionally active chromatin.

Differences at positions 31 and 87 are present in other forms of histone H3. The emergence of replacement histone H3.3 of *Tetrahymena* (Fig. 1, #19) coincides with a change from alanine to valine at position 31 and with the change from serine to glutamine at position 87. Barring other effects of the more divergent sequence of histone H3 in this protist, this correlation strengthens the idea that changes at these two positions affect the ability to a histone H3 protein to participate in the assembly of a nucleosome on DNA in non–S phase cells. In other words, they confer the functionality of a replacement histone.

Changes of residues 31 and 87 are inferred for two additional groups of species, for chlorophyta and for fungi (Fig. 6). In *Chlamydomonas,* only a single form of histone H3 exists (Waterborg et al. 1995). It has the same change at position 87 as in replacement H3.3 histone of *Tetrahymena* (Fig. 6). Experimental analysis is in progress to assess the ability of this histone to participate in the assembly of nucleosomes in non–S phase cells. Preliminary observations suggest that this is indeed the case. The basal level of histone H3 protein synthesis outside of S phase is quite significant. In addition, in light–dark synchronized cells, a burst of histone synthesis occurs when the dark period ends. Histone H3, newly synthesized under either condition, is incorporated into chromatin in the absence of DNA replication (Waterborg et al. 1995). Thus, the single histone H3 form of *Chlamydomonas* has the essential functional characteristic of a replacement histone H3.

In fungi, only residue 31 is changed from alanine to serine, and this change has occurred only in a limited number of species (Fig. 6). It is unknown whether this change has any effect on the abilities of fungal histone H3 to participate in the assembly of nucleosomes in non–S phase cells. However, it is noteworthy that the fungi that have a histone H3 with serine 31 support multicellular growth (Fig. 6).

Discussion

What makes a histone a functional replacement histone? One obvious requirement is that the histone protein must be available to participate in the assembly of a nucleosome in a cell which is not synthesizing DNA. One school of thought is that any histone gene, if constitu-

tively active or expressed in noncycling cells, would *de facto* become a replacement histone. It would be available to compete with histones scavenged from displaced nucleosomes to reassemble nucleosomes and to maintain or repair the nucleosomal packaging of DNA. In this model, the replacement histone H3 of animals (sequences 4–5), plants (sequence 10), and *Tetrahymena* (sequence 19) is recognized as a functional replacement histone solely because it is produced constitutively and has a sequence distinct from the replication H3 proteins in these species (sequences 1–3, 11–12, and 18, respectively) (Fig. 1). Inherent in this model is the expectation that histone H3 from displaced nucleosomes can be reused efficiently and that it is only the availability of another H3 protein form that leads to its gradual incorporation into the transcriptionally active chromatin of terminally differentiated cells (Zweidler 1984).

This model states that none of the sequence differences between replication and replacement histone H3 forms are of any consequence (Thatcher et al. 1994). It would predict that loss of the cell cycle control of a replication-dependent H3 gene would create a functional replacement histone gene. It would not envision a barrier against the sequence homogenization and gene conversion that has been demonstrated among replication-dependent histone genes (Wells et al. 1989). Only in sea urchins is it possible that gene conversion between replication and replacement H3 genes at position 81 has occurred (Fig. 6, #3 and #5), but this did not involve residues that may play a role in the ability of histone H3 to participate in the assembly of nucleosomes in a non–S phase cell. In fact, the sequence of the replacement histone H3 in animals and in plants has remained unchanged during speciation while changes did occur in the replication-dependent histone H3 genes (Fig. 6). No instance is known in which a replication H3 variant protein is produced constitutively or in which a replacement H3 protein form is made in a replication-dependent manner. These observations support the notion that the amino acid changes, associated with the creation of intron-bearing replacement histone H3 genes in animals and in plants, are functionally important. Experiments are in progress to test this conclusion.

References

Arents G, Moudrianakis EN (1993) Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. Proc Natl Acad Sci USA 90:10489–10493.

Arents G, Moudrianakis EN (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc Natl Acad Sci USA 92:11170–11174

Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN (1991) The nucleosomal core histone octamer at 3.1 Å resolution. A tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci USA 88:10138–10148

Bannon GA, Calzone FJ, Bowen JK, Allis CD, Gorovsky MA (1983) Multiple, independently regulated, polyadenylated messages for histone H3 and H4 in Tetrahymena. Nucleic Acids Res 11:3903–3917

Baxevanis AD, Landsman D (1996) Histone sequence database: a compilation of highly-conserved nucleoprotein sequences. Nucleic Acids Res 24:245–247

Biggin MD, Tjian R (1988) Transcription factors that activate the ultrabithorax promoter in developmentally staged extracts. Cell 53:699–711

Björkroth B, Ericsson C, Lamb MM, Daneholt B (1988) Structure of the chromatin axis during transcription. Chromosoma 96:333–340

Brocard MP, Rousseau D, Lawrence JJ, Khochbin S (1994) Two mRNA species encoding the differentiation-associated histone H1° are produced by alternative polyadenylation in mouse. Eur J Biochem 221:421–425

Chaboute ME, Chaubet N, Clement B, Gigot C, Philipps G (1988) Polyadenylation of histone H3 and H4 mRNAs in dicotyledonous plants. Gene 71:217–223

Chaboute ME, Chaubet N, Gigot C, Philipps G (1993) Histones and histone genes in higher plants: structure and genomic organization. Biochimie 75:523–531

Chaubet N, Chaboute ME, Clement B, Ehling M, Philipps G, Gigot C (1988) The histone H3 and H4 messenger RNAs are polyadenylated in maize. Nucleic Acids Res 16:1295–1304

Chaubet N, Clement B, Gigot C (1992) Genes encoding a histone H3.3-like variant in Arabidopsis contain intervening sequences. J Mol Biol 225:569–574

Desjardins E, Hay N (1993) Repeated CT elements bound by zinc-finger proteins control the absolute and relative activities of the two principal human c-myc promoters. Mol Cell Biol 13:5710–5724

Fabry S, Müller K, Lindauer A, Park PB, Cornelius T, Schmitt R (1995) The organization, structure and regulatory elements of Chlamydomonas histone genes reveal features linking plant and animal genes. Curr Genet 28:333–345

Fahrner K, Yarger J, Hereford L (1980) Yeast histone mRNA is polyadenylated. Nucleic Acids Res 8:5725–5737

Fodinger M, Ortner S, Plaimauer B, Wiedermann G, Scheiner O, Duchene M (1993) Pathogenic Entamoeba histolytica: cDNA cloning of a histone H3 with a divergent primary structure. Mol Biochem Parasitol 59:315–322

Heintz N (1991) The regulation of histone gene expression during the cell cycle. Biochim Biophys Acta 1088:327–340

Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenies. Science 264:671–677

Kapros T, Bögre L, Németh K, Bakó L, Györgyey J, Wu S-C, Dudits D (1992) Differential expression of histone H3 gene variants during cell cycle and somatic embryogenesis in alfalfa. Plant Physiol 98:621–625

Kapros T, Robertson AJ, Waterborg JH (1995) Histone H3 transcript stability in alfalfa. Plant Mol Biol 28:901–914

Kirsh AL, Groudine M, Challoner PB (1989) Polyadenylation and U7 snRNP-mediated cleavage—alternative modes of RNA 3' processing in two avian histone H1 genes. Genes Dev 3:2172–2179

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania University, University Park, PA

Laner M, de Bruin D, Wertheimer SP, Ravetch JV (1994) Transcriptional and nucleosomal characterization of a subtelomeric gene cluster flanking a site of chromosomal rearrangements in Plasmodium falciparum. Nucleic Acids Res 22:4176–4182

Li B, Adams CC, Workman JL (1994) Nucleosome binding by the constitutive transcription factor Sp1. J Biol Chem 269:7756–7763

Lu Q, Wallrath LL, Granok H, Elgin SC (1993) (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene. Mol Cell Biol 13:2802–2814

Mannironi C, Bonner WM, Hatch CL (1989) H2A.X, a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals. Nucleic Acids Res 17:9113–9126

Marzluff WF (1989) Histone gene expression not coupled to DNA synthesis. In: Hnilica LS, Stein GS, Stein JL (eds) Histones and other basic proteins. CRC Press, Boca Raton, FL

Marzluff WF (1992) Histone 3' ends: essential and regulatory functions. Gene Expr 2:93–97

Matthews HR, Waterborg JH (1985) Reversible modifications of nuclear proteins and their significance. In: Freedman RB, Hawkins HC (eds) The enzymology of post-translational modification of proteins, vol 2. Academic Press, London

Michalon P, Couturier R, Bender K, Hecker H, Marion C (1993) Structural analysis of Trypanosoma brucei brucei chromatin by limited proteolysis. An electrical birefringence study. Eur J Biochem 216:387–394

Muller SB, Rensing SA, Maier UG (1994) The cryptomonas histone H4-encoding gene: structure and chromosomal location. Gene 150:299–302

O'Brien T, Wilkins RC, Giardina C, Lis JT (1995) Distribution of GAGA protein on Drosophila genes in vivo. Genes Dev 9:1098–1110

O'Donnell KH, Wensink PC (1994) GAGA factor and TBF1 bind DNA elements that direct ubiquitous transcription of the α1-tubulin gene. Nucleic Acids Res 22:4712–4718

O'Donnell KH, Chen CT, Wensink P (1994) Insulating DNA directs ubiquitous transcription of the Drosophila melanogaster α1-tubulin gene. Mol Cell Biol 14:6398–6408

Old RW, Woodland HR (1984) Histone genes: not so simple after all. Cell 38:624–626

Osley MA (1991) The regulation of histone synthesis in the cell cycle. Annu Rev Biochem 60:827–861

Peltz SW, Brewer G, Bernstein P, Hart PA, Ross J (1991) Regulation of mRNA turnover in eukaryotic cells. Crit Rev Eukaryot Gene Expr 1:99–126

Puerta C, Martin J, Alonso C, Lopez MC (1994) Isolation and characterization of the gene encoding histone H2A from Trypanosoma cruzi. Mol Biochem Parasitol 64:1–10

Robertson AJ, Kapros T, Dudits D, Waterborg JH (1995) Identification of the three highly expressed replacement histone H3 genes of alfalfa. DNA Seq (in press)

Sadler LA, Brunk CF (1992) Phylogenetic relationships and unusual diversity in histone H4 proteins within the Tetrahymena pyriformis complex. Mol Biol Evol 9:70–84

Saint-Guily A, Schantz ML, Schantz R (1994) Structure and expression of a cDNA encoding a histone H2A from Euglena gracilis. Plant Mol Biol 24:941–948

Schlimme W, Burri M, Bender K, Betschart B, Hecker H (1993) Trypanosoma brucei brucei: differences in the nuclear chromatin of bloodstream forms and procyclic culture forms. Parasitology 107:237–247

Soto M, Requena JM, Jimenez-Ruiz A, Alonso C (1991) The messenger RNA coding for the nucleosomal protein H2A of Leishmania is polyadenylated and has stem-loops at the 3' end. Nucleic Acids Res 19:4554

Soto M, Requena JM, Morales G, Alonso C (1994) The Leishmania infantum histone H3 possesses an extremely divergent N-terminal domain. Biochim Biophys Acta 1219:533–535

Standart N, Dale M (1993) Regulated polyadenylation of clam maternal mRNAs in vitro. Dev Genet 14:492–499

Swofford DL (1993) PAUP: Phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Campaign, IL

Thatcher TH, MacGaffey J, Bowen J, Horowitz S, Shapiro DL, Gorovsky MA (1994) Independent evolutionary origin of histone H3.3-like variants of animals and Tetrahymena. Nucleic Acids Res 22:180–186

Tsukiyama T, Becker PB, Wu C (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 367:525–532

Urban MK, Zweidler A (1983) Changes in nucleosomal core histone variants during chicken development and maturation. Dev Biol 95:421–428

VanDaal A, White EM, Elgin SC, Gorovsky MA (1990) Conservation of intron position indicates separation of major and variant H2As is an early event in the evolution of eukaryotes. J Mol Evol 30:449–455

Vanfleteren JR, VanBun SM, VanBeeumen JJ (1989) The histones of Caenorhabditis elegans. No evidence of stage specific isoforms—an overview. FEBS Lett 257:233–237

Van Holde KE (1989) Chromatin. In: Chromatin. Springer Verlag, New York

Wall G, Varga-Weisz PD, Sandaltzopoulos R, Becker PB (1995) Chromatin remodeling by GAGA factor and heat shock factor at the hypersensitive Drosophila hsp26 promoter in vitro. EMBO J 14:1727–1736

Waterborg JH (1990) Sequence analysis of acetylation and methylation of two histone H3 variants of alfalfa. J Biol Chem 265:17157–17161

Waterborg JH (1991) Multiplicity of histone H3 variants in wheat, barley, rice and maize. Plant Physiol 96:453–458

Waterborg JH (1992) Existence of two histone H3 variants in dicotyledonous plants and correlation between their acetylation and plant genome size. Plant Mol Biol 18:181–187

Waterborg JH (1993) Histone synthesis and turnover in alfalfa. Fast loss of highly acetylated replacement histone H3.2. J Biol Chem 268:4912–4917

Waterborg JH, Robertson AJ, Tatar DL, Borza CM, Davie JR (1995) Histones of Chlamydomonas reinhardtii. Synthesis, acetylation and methylation. Plant Physiol 109:393–407

Wells JRE, Coles LS, Robins AJ (1989) Organization of histone genes and their variants. In: Hnilica LS, Stein GS, Stein JL (eds) Histones and other basic proteins. CRC Press, Boca Raton, FL

Wilhelm ME, Toublan B, Fujita RA, Wilhelm FX (1988) Histone H4 messenger RNA is stored as a small cytoplasmic rnp during the G2 phase in Physarum polycephalum. Biochem Biophys Res Commun 153:162–171

Williams AS, Ingledue TC, Kay BK, Marzluff WF (1994) Changes in the stem-loop at the 3' terminus of histone mRNA affect its nucleoplasmic transport and cytoplasmic regulation. Nucleic Acids Res 22:4660–4666

Wu S-C, Györgyey J, Dudits D (1989) Polyadenylated H3 histone transcripts and H3 histone variants in alfalfa. Nucleic Acids Res 17:3057–3063

Zhong R, Roeder RG, Heintz N (1983) The primary structure and expression of four cloned human histone genes. Nucleic Acids Res 11:7409–7425

Zweidler A (1984) Core histone variants of the mouse: primary structure and differential expression. In: Stein GS, Stein JL, Marzluff WF (eds) Histone genes: structure, organization and regulation. John Wiley & Sons, New York