REVIEW

nature structural biology

Phylogenomics of the nucleosome

Harmit S Malik[1] & Steven Henikoff[1,2]

Histones are best known as the architectural proteins that package the DNA of eukaryotic organisms, forming octameric nucleosome cores that the double helix wraps tightly around. Although histones have traditionally been viewed as slowly evolving scaffold proteins that lack diversification beyond their abundant tail modifications, recent studies have revealed that variant histones have evolved for diverse functions. H2A and H3 variants have diversified to assume roles in epigenetic silencing, gene expression and centromere function. Such diversification of histone variants and 'deviants' contradicts the perception of histones as monotonous members of multigene families that indiscriminately package and compact the genome. How these diverse functions have evolved from ancestral forms can be addressed by applying phylogenetic tools to increasingly abundant sequence data.

It has been fifty years since the elucidation of the structure of DNA heralded the beginning of molecular biology. But more recent scientific attention has shifted to the other stoichiometric component of chromosomes, architectural proteins that facilitate the organization, replication, expression and compaction of genes in all cells. The diversity of form and function of these proteins is responsible for dictating when genetic information is switched on or off, how it is faithfully transmitted through cell division and even how it is repaired. This interplay between form and function is particularly striking in the case of the four core histone proteins that carry out this role in virtually all eukaryotes. The major aim of this review is to summarize the ancient evolutionary origins of this quartet of proteins, as well as more recent specializations that have further articulated their roles in gene expression and chromosome segregation. Phylogenetic comparisons resolve functionally important distinctions that are too subtle for structural comparisons to delineate. This approach can be generally applied to highly conserved proteins.

Archaeal histones

It is now apparent that the evolutionary origins of eukaryotic histones can be traced back to the archaeal histones first discovered in *Methanothermus fervidus* (*Methanobacterium clade*)[1] and now known to be present in almost all lineages of the Euryarchaeota (**Fig. 1a**)[2]. Most archaeal histones comprise a single histone fold domain (HFD), characterized by three α-helices, and two intervening loops, with no N-terminal or C-terminal 'tails.' The genome of *M. fervidus* encodes at least two distinct histones, HMfA (for histone *M. fervidus* A) and HMfB, which have been shown to compact DNA. A tetramer of these histone proteins is able to protect 60 base pairs of packaged DNA from nuclease digestion, suggesting a single wrap of DNA around the tetramer[3]. Like eukaryotic nucleosomes, this tetramer is able to induce supercoiling upon packaging closed circular DNA[4]. HMfA and HMfB form both homodimers and heterodimers. Because of differences in their amino acid composition, the different tetramers are predicted to have altered biological properties. Consistent with this view, HMfA predominates in exponentially growing cells, whereas HMfB is abundant in stationary cells where it is predicted to increase the compaction of genomic DNA[5]. Notably, each histone is believed to participate equally in either distinguishable position of the two-fold symmetric archaeal tetramer[2].

Not all euryarchaeal lineages have a similar complement of two histones. The histone gene complement in completely sequenced euryarchaeal genomes varies from one to six genes (**Fig. 1a**), greatly affecting the potential tetramer combinations that could form. A phylogenetic analysis of the archaeal histones reveals that the duplicate genes found in the same archaeal genome typically group close together (**Fig. 1b**), suggesting that recent duplications rather than ancient paralogy are responsible for archaeal histone gene complements. Surprisingly, both completed genomes from Thermoplasma lack histone genes entirely. They are unique among the Archaea (so far) to possess eubacterial HU-like proteins[2] suggesting that the acquisition of HU proteins, most likely by horizontal gene transfer, led to the subsequent loss of histone genes from Thermoplasma.

A major advance in understanding the evolutionary history of the eukaryotic nucleosome came from the discovery of an unusual 'doublet' histone, first in *Methanopyrus kandleri*[6] and subsequently in *Halobacterium* species NRC1 (ref. 7) (blue lineages in **Fig. 1a,b**). These histones are twice as long as typical archaeal histones and consist of an end-to-end duplication of the histone fold. An end-to-end duplication is evolutionarily significant because unlike stand-alone histone proteins, end-to-end duplications are no longer required to participate in all positions of the histone tetramer; the N-terminal and C-terminal histone domains are able to subfunctionalize. This subfunctionalization is apparent in a comparison of the primary sequence of the various archaeal histone domains. We compared a consensus derived from all the predicted 'singlet' histone genes (presented in Logos format) to the four instances of histone domains from the two doublet genes

[1]Basic Sciences Division and [2]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. Correspondence should be addressed to S.H. (steveh@fhcrc.org).

Published online 28 October 2003, doi:10.1038/nsb996

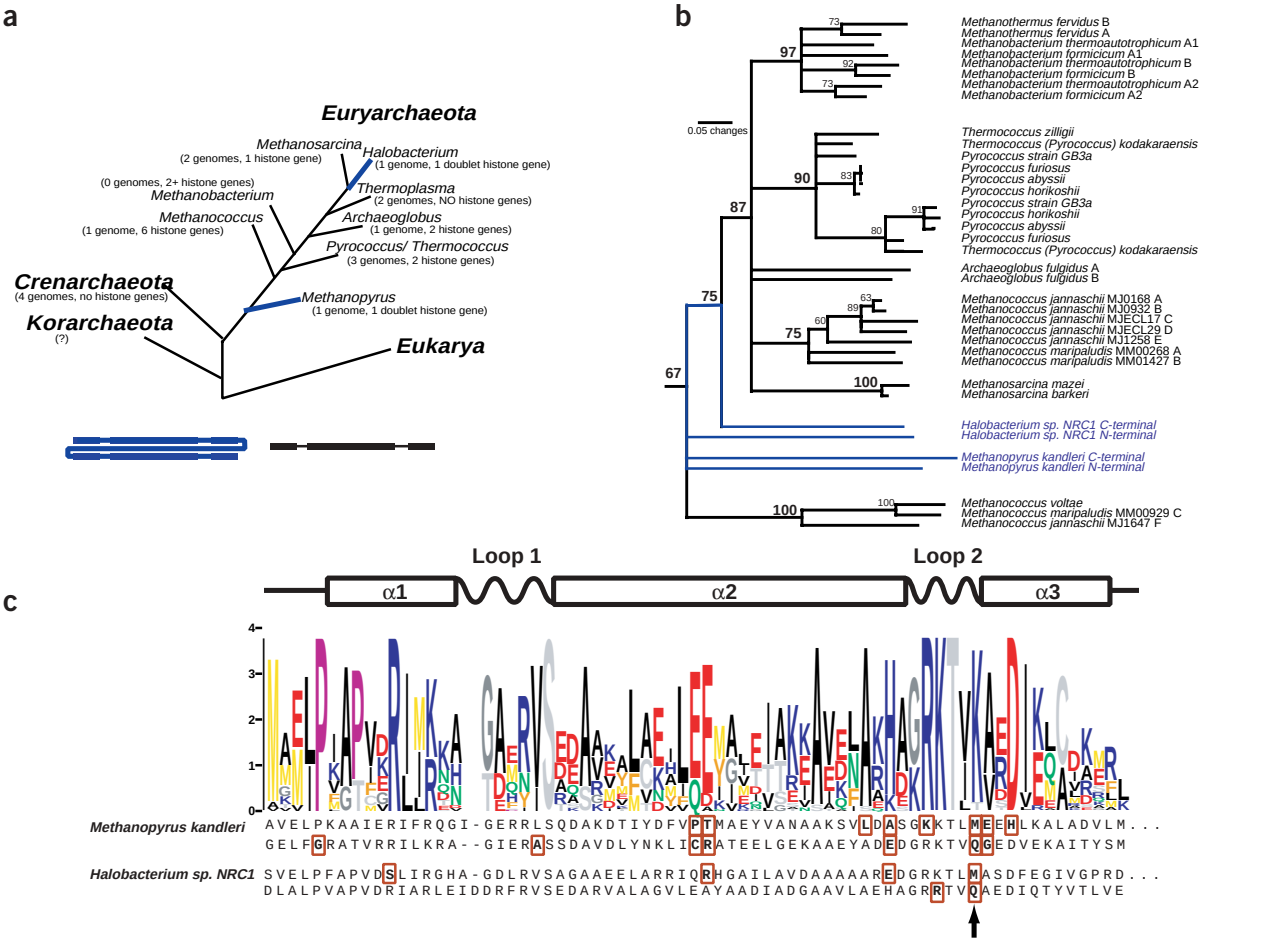

Figure 1 Archaeal histones. (**a**) A schematic archaeal phylogeny concentrating on the Euryarchaeota, showing the number of available completely sequenced genomes, and the number of histone genes identified in them. (**b**) A neighbor-joining tree of the various archaeal histone domains. Nodes with bootstrap support >50% are shown, whereas other nodes are collapsed. In the case of the 'doublet' histones, both the N- and C-terminal domains are represented separately (blue lineages). (**c**) An alignment of the singlet histone domains is shown in Logos format (where the tallest residues are invariant) and compared with the doublet histone N- and C- terminal histone domains below. Several residues in the doublet histones show evidence of marked deviation from the singlet histone consensus (bold residues, red boxes). In one position (arrow), the same conserved lysine residue has changed to a methionine and glutamine in N- and C-terminal histones, respectively.

(**Fig. 1c**). It is clear that the evolutionary constraint apparent in the singlets is relaxed in certain domains of the doublets, in particular in the middle of α-helix 2 and surrounding loop 2 (shown in bold with red boxes in **Fig. 1c**). It is expected that subfunctionalization would lead to relaxed constraints in some parts of the doublet histones. However, one position is particularly interesting because a well conserved lysine in the singlet histones has changed to methionine and glutamine in the N- and C- terminal histone folds of both the doublet histone genes (arrow in **Fig. 1c**). Their evolutionary divergence (**Fig. 1b**) suggests that this is not because of recent ancestry. Instead, this could reflect a common constraint imposed on both doublets consistent with their newly specialized positions in the tetrameric structure. As more doublet histones are discovered, this change and its functional relevance might be further validated biochemically and structurally.

Archaeal histones to eukaryotic histones

The doublet histone was probably an important and perhaps essential intermediate in the transition from archaeal to eukaryotic histones[8]. The constraints imposed on the single histone fold might have been too great to allow the diversification seen in the eukaryotic histone octamer. However, the formation of the doublet histone directly resulted in an asymmetric dimer that could have preceded the actual separation of the H3-H4 and H2A-H2B genes (**Fig. 2a**). Structures are available for the histone proteins from *M. fervidus*, *M. kandleri* and a few eukaryotes[8–11]. These reveal a high degree of structural similarity between the various archaeal and eukaryotic histones (**Fig. 2b–g**). The *M. kandleri* protein, which contains two histone domains, dimerizes via the C-terminal helices of each N-terminal histone fold. The N-terminal histone fold in the end-to-end dimer thus corresponds structurally to H3, and the C-terminal fold to H4 (ref. 8). The HMk (histone *M. kandleri*) protein aligns well with the other histone dimers, with an r.m.s. deviation of 1.5–2.7 Å (ref. 8). Notably, one of the conserved differences in the alignment between singlet and doublet archaeal histones is the change from lysine to methionine in the N-terminal domain described above; this same position is highly constrained as a methionine only in the orthologous H3 lineage. Given the long time that has elapsed, it is difficult to determine whether this

reflects common ancestry or convergent evolution, but nonetheless strongly suggests a common constraint at the interface between the two H3–H4 dimers and the HMk dimer.

What was the order of events leading up to the eukaryotic complement of four histones in an octamer? Several phylogenetic treatments have suggested an early origin of H4 or H2A[6,12]. However, the high structural homology between the different histones complicates attempts at aligning them, because the introduction of gaps can artifactually skew the phylogenetic result. Given the different selective constraints on each histone backbone, it is difficult to separate phylogenetic history from subsequent specialization events. Nonetheless, the ability to form a tetramer must have been ancestrally conserved, as this is a common feature of both archaeal and eukaryotic histones. This suggests that H3 and H4 evolved before H2A and H2B. Among eukaryotic core histones, only H2A and H3 self-dimerize in the histone octamer[10,11]. The H3 self-dimerization domain is reminiscent of that seen for the HMk archaeal doublet histone[8], whereas the H2A self-dimerization domain is unusual and might be too weak to have been an ancestrally retained dimerization domain. Indeed, under physiological conditions H3 and H4 are found only as heterotypic tetramers, whereas H2A and H2B are found only as heterodimers[13]. Under the doublet intermediate model (above), both H3 and H4 would have had a simultaneous origin, followed by a second specialization of a doublet to give rise to H2A and H2B.

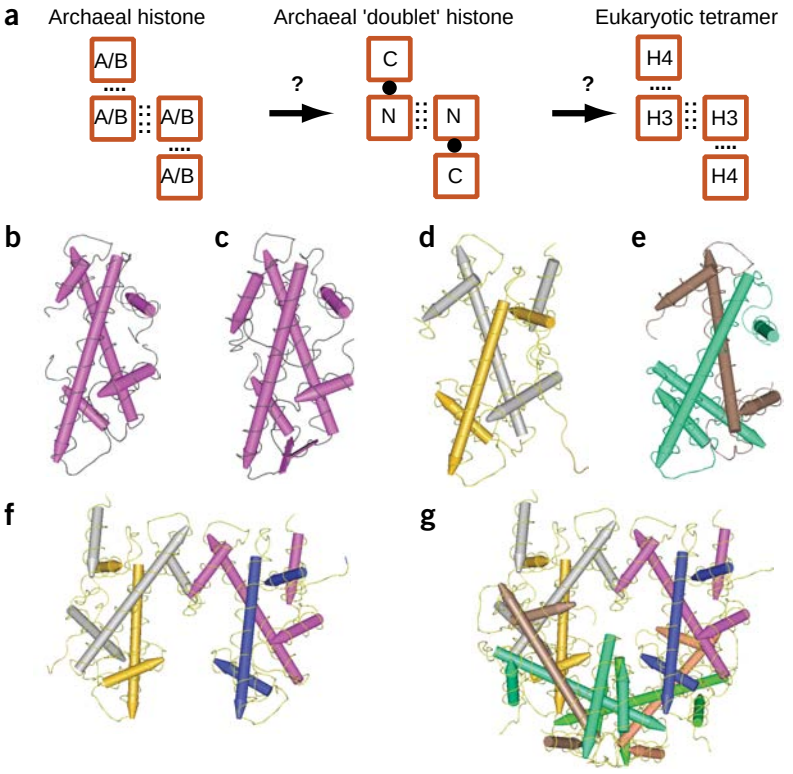

Figure 2 Archaeal to eukaryotic histones. (**a**) A proposed transition from archaeal to eukaryotic histone tetramers involving an intermediate stage of doublet histones, similar to those found in *M. kandleri*[6]. This intermediate may have led to the evolution of a H3-H4 ancestral doublet, followed by a gene split to give rise to H3 and H4 (ref. 8). H2A and H2B may have arisen from a separate evolutionary event involving a second doublet histone (not shown). (**b**–**e**) Crystal structures of various archaeal and eukaryotic histone dimers. These include (**b**) a homodimer of HMfA *from M. fervidus* (PDB entry 1B67)[9],(**c**) the doublet histone from *M. kandleri* (PDB entry 1F1E) [8] and (**d**) H3–H4 and (**e**) H2A–H2B heterodimers from *Xenopus laevis* (PDB entry 1AO1)[10]. Note the presence of an extra N-terminal α-helix in H3. A C-terminal α-helix in H2B is not shown for clarity of presentation. (**f**) The two H3–H4 dimers organize along the C-terminal end of H3 to result in the eukaryotic histone tetramer[10]. (**g**) The addition of two H2A–H2B heterodimers results in the histone octamer characteristic of eukaryotes[10,11,13]. All structure presentations were made using Cn3D (version 4.11; available from NCBI at http://www.ncbi.nlm.nih.gov/structure/CN3D/Cn3d.shtml).

Clearly, the origin of H2A–H2B dimers represents a major shift in strategy from tetramers with nearly one wrap of DNA to octamers with nearly two wraps. It has been suggested that the increasing genome size of the ancestral eukaryote has selected for a greater degree of packaging afforded by the octamer but not the tetramer[2]. This might not simply have been a stoichiometric event (the size of the archaeal *Methanosarcina acetivorans* genome is 5.7 Mb, whereas those of eukaryotes *Giardia lamblia* and *Encephalitozoon cuniculi* are 12 Mb and 2.7 Mb, respectively). The high degree of DNA compaction and rapid rate of condensation required for eukaryotic-specific mitosis might also have selected for this transition. Other means of achieving greater compaction would have been the acquisition of histone tails and linker histones such as H1. Once present, these tails and linker histones would impede access to genomic DNA, which might have led to them being heavily modified. These modifications would in turn have imposed a strict evolutionary constraint on the amino acid residues in the histone tails, which are essentially invariant, especially in H3 and H4 (see below)[14]. A high degree of compaction and speed of condensation would be an unnecessary cost to an archaeon that lacked mitosis and possessed few gene-poor regions, but may have been essential for the burgeoning, mitotic eukaryotic genome.

H3 variants but an invariant H4

The $[H3–H4]_2$ tetramer represents a direct structural and perhaps evolutionary link to the archaeal histone tetramer (**Fig. 2a–c,f**). Despite the fact that canonical H3 and H4 histones are among the most slowly evolving eukaryotic proteins, they differ substantially in their evolutionary specializations. Distinct variants of H3 have been usurped for special roles in transcription and even chromosome segregation, whereas H4's role has remained constant throughout eukaryotic evolution.

Major lineages of histones are distinguished from histone variants by both their timing of expression and their pattern of incorporation into genomes. The bulk of histones in eukaryotic cells are S-phase specific—that is, they are deposited during DNA replication to fill in the resulting gaps in nucleosomal arrays. In most metazoans, these S-phase-specific genes are present as multigene families, with each gene cluster typically encoding all four core histones and the non-HFD linker histone H1. S-phase-specific histones are highly expressed but only in a short burst coincident with DNA replication; they participate

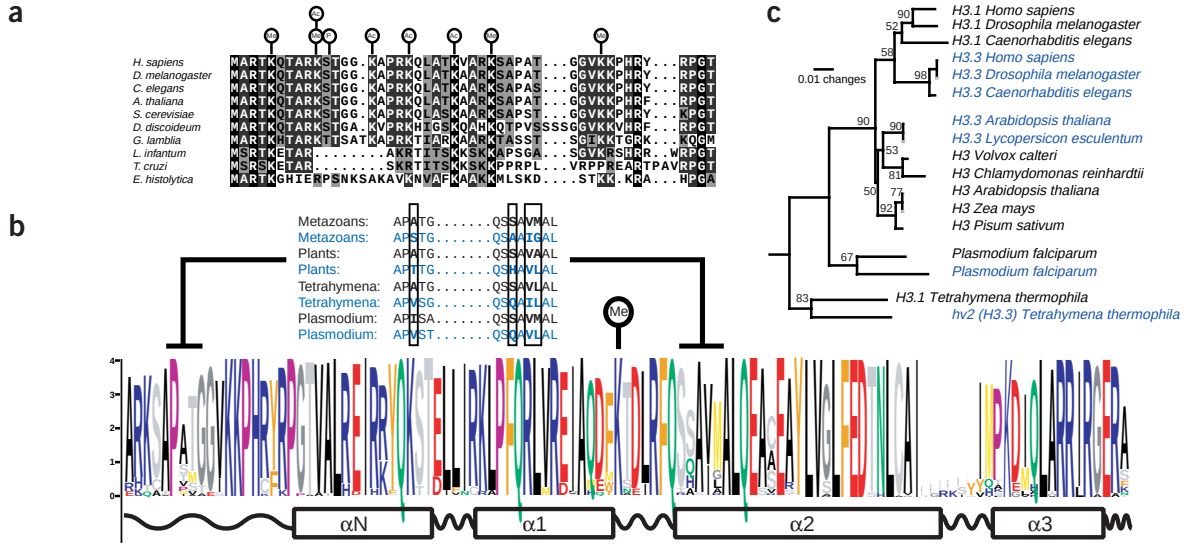

Figure 3 Canonical H3 histones and the H3.3 variant. (**a**) N-terminal tails of H3 and H3.3, with modifications shown above (Me, methylation; Ac, acetylation; P, phosphorylation)[14]. (**b**) A combined alignment in Logos format for the histone fold domains of H3 and H3.3. Except for four positions, the H3 and H3.3 proteins are practically identical within species. These four positions are shown above the Logos alignment derived from plants, animals, ciliates and apicomplexans. Also indicated is a conserved modification (Me) within the HFD. (**c**) A phylogeny of H3 and H3.3 from these four lineages strongly suggests independent evolutionary origins of H3.3 (refs. 17,18).

in the bulk, indiscriminate packaging of the eukaryotic genome. In contrast, 'orphan' histones are found outside gene clusters and are usually synthesized independent of S phase. As a result, whereas S-phase histones predominate in rapidly dividing cells, S-phase-independent histones continue to accumulate after DNA replication.

From a structural perspective, H3 has a key role in organizing the nucleosome. The two-fold symmetry of the eukaryotic nucleosome is organized along the dimerization interface of the two H3 molecules using their C-terminal ends, an example of a four-helix bundle[10,11]. Apart from its heterodimerization with H4, H3 also makes contacts with H2A and has at least two segments of specific contact with the nucleosomal DNA: just upstream of the αN helix where the H3 tail enters the nucleosome between the DNA gyres, and in the loop 1 region. At least two H3 variants are found in most eukaryotic lineages.

H3.3 is a replacement H3 variant that is encoded by 'orphan' genes devoid of the strict transcriptional and post-transcriptional controls to which canonical histone genes are subject. Thus, although H3.3 expression is heavily diluted by the S-phase-specific histones during DNA replication, it is the major H3 molecule that is available for deposition outside of S phase[15]. For instance, in long-lived neuronal cells, H3.3 is the dominant entity rather than the S-phase-specific H3 (ref. 16). These constitutively expressed histones are believed to be crucial for replacing nucleosomes that are lost during cellular processes such as transcription and DNA repair. As a consequence, replacement histones have been implicated in facilitating different states of chromatin.

H3.3 is distinguished from canonical H3 by differences at only a few positions. The N-terminal tails of H3 and H3.3 are nearly identical (**Fig. 3a**) and for the most part, these two histones show a remarkably high degree of evolutionary constraint throughout their length (**Fig. 3b**). However, from an evolutionary standpoint, these are not separate lineages. Indeed, it seems that a distinction between H3 and H3.3 types has arisen numerous times. Four such instances are evident in ciliates, apicomplexans, animals and plants (**Fig. 3c**)[17,18]. Notably, each of these specializations requires changes predominantly at only four positions, one in the N-terminal tail and three in the HFD α2 helix. Strictly speaking, this is not convergent evolution because the H3.3 versions do not all have the same amino acid residues in these four positions. However, it strongly suggests that a similar constraint has led to these repeated origins of distinguishable H3 and H3.3 types.

In Ascomycetes (fungi) such as *Saccharomyces cerevisiae*, only one type of gene is present. Comparison with the basally branching Basidiomycetes, which have both H3 and H3.3, led to the surprising conclusion that only the H3.3 version has been retained in Ascomycetes, presumably because H3.3 can deposit both during and after replication, whereas H3 cannot[19]. Indeed, we know of no eukaryotic genome that has been characterized with only H3 and not H3.3. Although there is little doubt regarding multiple origins of H3.3, it is certainly conceivable that ancestrally, an H3.3 was present in a small, predominantly transcriptionally active genome. The rapid expansion of eukaryotic genomes, large portions of which became silent in differentiated cells, may have selected for H3 and its expansion, both for increasing bulk packaging duties and to ensure transcriptional silencing where appropriate. In this regard, it is important to note that even in the archaeon *M. fervidus*, this division of labor is evident between HMfA and HMfB[5].

What is the functional consequence of the four amino acid substitutions in H3.3? The amino acid sequence of hv2 (H3.3) in *Tetrahymena thermophila* seems to be less important than its constitutive expression pattern[20]. However, in *Drosophila melanogaster*, three amino acid differences clustered in the HFD are critical for replication-independent deposition[19]. Based on the crystal structure, these three residues seem to be solvent accessible. Whether these affect the deposition of nucleosomes or their strength of packing still remains to be determined, but the correlation between H3.3 deposition and transcriptional activity *in vivo* strongly suggests it is more transcriptionally permissive, and may function akin to H2A.Z (see below).

The eukaryotic genome is predominantly packaged into chromatin containing histones, but in many male pronuclei, a specialized

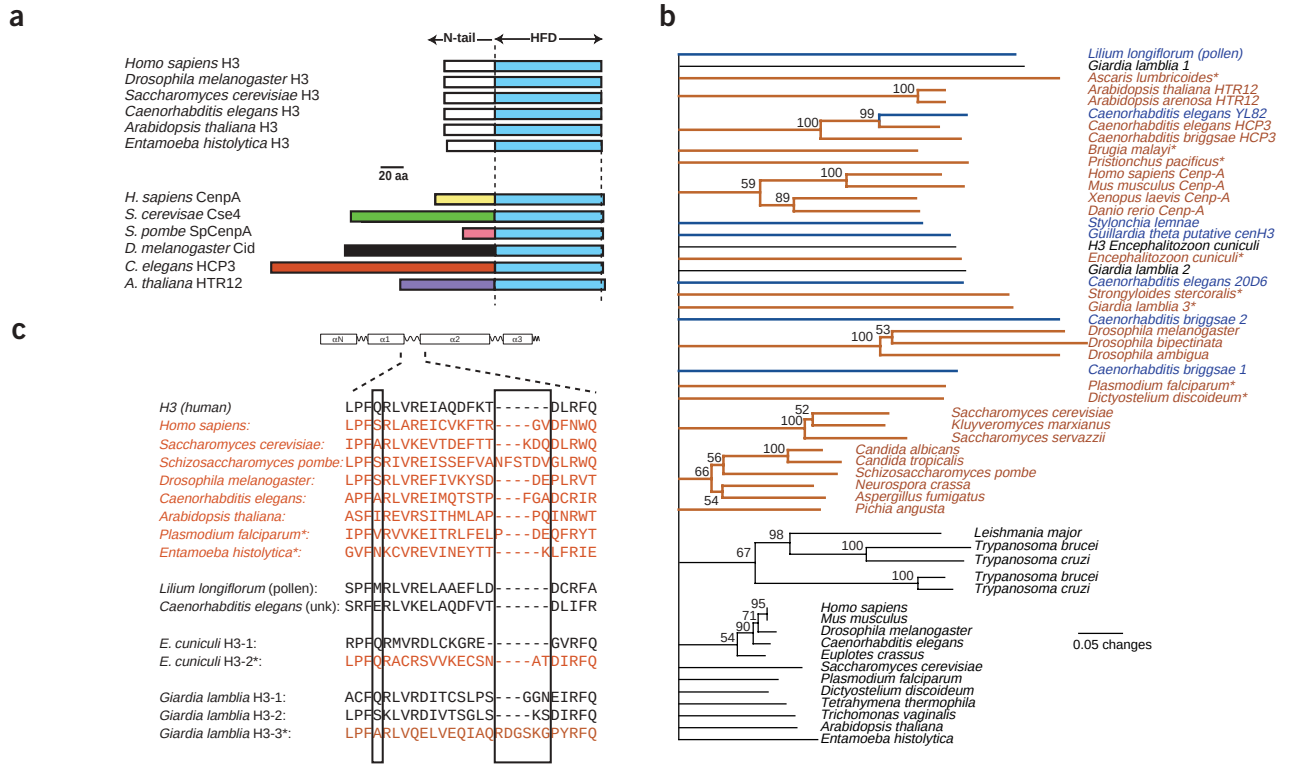

Figure 4 Centromeric H3 variants. (**a**) Schematic of canonical H3 (and H3.3) compared with centromeric H3 lineages (CenH3s). The N-terminal tails of the former set are nearly invariant (**Fig. 3a**) but those from CenH3s vary greatly in size and sequence[25]. (**b**) A neighbor-joining phylogeny of the HFD from canonical and CenH3s shows the dichotomy between these two groups. Canonical H3s (black) typically have short branches except in some genomes, whereas CenH3s (orange lineages) have long branches. In addition, divergent orphan H3 lineages (blue) are also found. All branches with bootstrap support <50% are collapsed. (**c**) Bioinformatic criteria to identify CenH3s include a longer, more divergent N-terminal tail as well as lack of a conserved glutamine and a longer loop 1 in the HFD[25]. These criteria have led to the identification of CenH3s in a variety of lineages and tentative assignments in some newly sequenced genomes (indicated with an asterisk).

chromatin structure is often found that is devoid of or severely deficient in histones. Instead, they are packaged in protamines, highly basic architectural proteins that might themselves have evolved from linker histones[21]. H3.3 might have a role in the transitioning from H3-containing chromatin to protamines[22]. Another H3 variant survives the protamine transition in mammals (and this survival led to its discovery): CenpA, the first example of a centromere-specific H3 (or CenH3)[23,24].

CenH3s are histone H3 variants that are specialized for packaging chromatin at eukaryotic centromeres. Eukaryotes are defined on the basis of having a nuclear envelope, but mitosis is a key process that sets them apart from prokaryotes. Mitosis depends on the presence of centromeres, sites of spindle attachment to eukaryotic chromosomes that ensure their correct segregation.

CenH3s are atypical histone H3 variants. Whereas the canonical H3 and H3.3 variants are well conserved throughout their length, CenH3s have no sequence similarity to H3 in their N-terminal tail, which can vary from 20 to ~200 amino acids in different lineages (**Fig. 4a**). In addition, CenH3s are only ~50% identical to H3s in the HFD[25]. A phylogeny consisting of representative canonical H3 (black lineages) and centromeric histones (orange lineages) highlights this dichotomy: whereas canonical H3s are highly constrained, centromeric histones are evolving rapidly (**Fig. 4b**). For instance, the amino acid divergence between canonical H3s in *Entamoeba histolytica* and human genomes, which shared a common ancestor close to a billion years ago, is less than that between CenH3s of two *Drosophila* species that diverged less than 15 million years ago[26]!

Only a handful of CenH3 lineages have been experimentally validated from fungi, animals and plants[27–32]. These have, in turn, led to bioinformatic criteria for identifying additional putative CenH3s. Thus, CenH3s have a divergent N-terminal tail, and typically lack a conserved glutamine residue in the α1 helix, but the most useful criterion is that all CenH3s seem to have a longer loop 1 region than their canonical counterparts (**Fig. 4c**). Loop 1 is one of the segments where H3 can make specific contacts with nucleosomal DNA[10,11] and a longer loop 1 region may lead to a more articulated contact and thereby to specificity[33,34]. The combination of these criteria unambiguously establishes the CenH3 gene in some genomes, and allows a tentative assignment of CenH3 function to one of a divergent set of H3 genes in other genomes. For instance, the two H3s from *E. cuniculi* are equally divergent, and have the conserved glutamine, but one gene is three amino acids longer in its loop 1 region than the other. Similarly, in *Giardia lamblia*, there are three H3 genes, all of which are very divergent in their HFDs. However, two of these have a more canonical N-terminal tail, whereas the third has a longer loop 1 region and is likely to be the CenH3.

Why are CenH3s evolving so rapidly? Pairwise comparisons between closely related plant and animal species reveal that in contrast

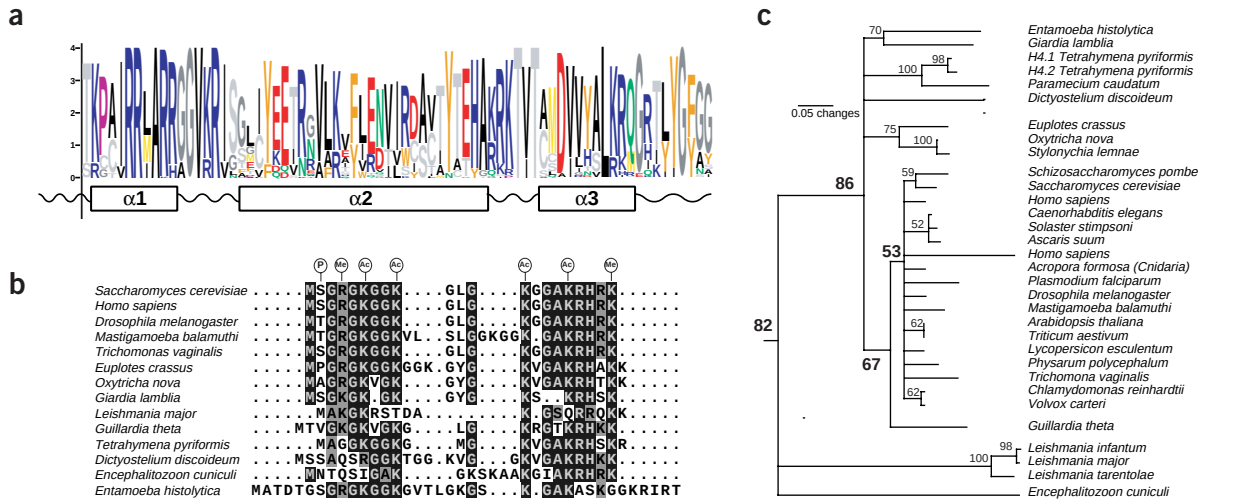

Figure 5 H4 histones. (**a**,**b**) A high degree of conservation of the H4 proteins is observed in both their (**a**) HFD and their (**b**) N-terminal tails where some residues are post-translationally modified[14]. (**c**) A neighbor-joining phylogeny of H4 proteins reveals the lack of any variant lineages outside of the canonical form.

to all conventional histones, CenH3 proteins are subject to positive selection[32,35]. This positive selection seems to be driven by altered DNA-binding specificity of the N-terminal tail and loop 1, which may be due to meiotic competition between different centromeric satellites for evolutionary dominance[25]. Chimeric swaps between the HFDs of different *Drosophila* species reveal that the loop 1 region is both necessary and sufficient for the centromeric localization of CenH3 proteins, confirming the adaptive evolution findings and validating the bioinformatic criterion (see above)[34]. Furthermore, the N-terminal tails harbor binding determinants for different proteins, including other kinetochore components[26,36]. The configurations of these components change between different lineages, which might explain the large differences in N-terminal tail sequence and size.

From a phylogenetic standpoint, no features clearly demarcate CenH3s as being a distinct lineage from canonical H3s. The phylogeny of the various H3s suffers from a remarkably poor lack of resolution (few nodes of strong bootstrap support). This suggests that the usual assumption of orthology of the CenH3s is not justified. Instead, an 'orphan' CenH3 may arise multiple times in the course of evolution and, because of rapid evolution, will quickly diverge from its canonical counterparts. For instance, even if two different CenH3s arise separately from H3s, the H3s will appear closer to each other in phylogenies because they are so evolutionarily constrained. Indeed, fungal CenH3 phylogeny is not congruent with fungal species phylogeny (**Fig. 4b**) and this invention of CenH3s seems to have occurred at least three times in fungal evolution (H.S.M., unpublished data).

Other orphan H3 lineages are also evident (**Fig. 4b**, indicated in blue). In all the histone phylogenies, outlier lineages typically fall into three categories: functionally specialized lineages, ancestral eukaryotic lineages that diverged early and recent lineages subject to relaxed selection. Relaxed selective constraints can account for the more rapid rate of histone evolution in Microsporidia, which were originally believed to have diverged early; however, with better phylogenetic methodology, it is clear that they represent a relatively young fungal lineage (such as *E. cuniculi*)[37].

H3 outlier lineages include a pollen-specific form from *Lilium longiflorum*[38] and at least four H3 genes from *Caenorhabditis elegans* and *Caenorhabditis briggsae* that are not orthologous. These outlier lineages appear as divergent as the CenH3 variants, but their function is still unknown. A meiosis-specific histone H3 gene in the ciliate *Euplotes crassus* (vH3_Ec) has also been implicated in the formation (or expression) of the macronucleus[39]. vH3_Ec seems to be recently derived from the canonical H3 lineage in *Euplotes crassus* for this ciliate-specific role. An H3 gene from the algal nucleomorph *Guillardia theta* has been identified as a putative CenH3 (ref. 40). However, this does not fit the bioinformatic criterion for a CenH3 (see above) and this genome, like that of *E. cuniculi*, has been subject to bottlenecks and relaxed selection; this gene may thus represent a divergent yet canonical H3.

Like canonical H3, histone H4 is one of the most slowly evolving proteins, with hardly any amino acid changes evident in fungi, plants and animals (**Fig. 5a**). A high degree of constraint is to be expected as H4 makes extensive protein-protein contacts with the other three histones: only some positions in α2 appear to tolerate any amino acid substitutions at all. The N-terminal tail of H4 is also modified extensively by kinases, acetyltransferases and methyltransferases. The modified residues and the rest of the N-terminal tail are highly constrained except in some outlier lineages (**Fig. 5b**), strongly suggesting an early origin and maintenance of histone modifications in the eukaryotic lineage. Some outlier lineages are the only notable feature in the phylogeny of representative H4 proteins (**Fig. 5c**). There are no sequence distinctions between S-phase-specific and replication-coupled H4 genes in plants and animals, and in the case of mammals, the two forms are produced as a result of alternative processing of the same primary transcript[27]. Thus, there is little evidence for any H4 functional specialization.

H2A variants but an invariant H2B

There are prominent parallels between the H2A–H2B and H3–H4 dimers. Like H3, the two H2A molecules interact with each other in the nucleosome, whereas H2B, like H4, does not. The H2B lineage, like H4, includes few variants, whereas H2A, like H3, has extensively specialized for myriad roles in transcription and DNA repair.

The interaction between the two H2A proteins in the nucleosome occurs via their loop 1 domains, an arrangement not seen for either H3 self-dimerization or archaeal histones[10,11]. H2A is characterized by

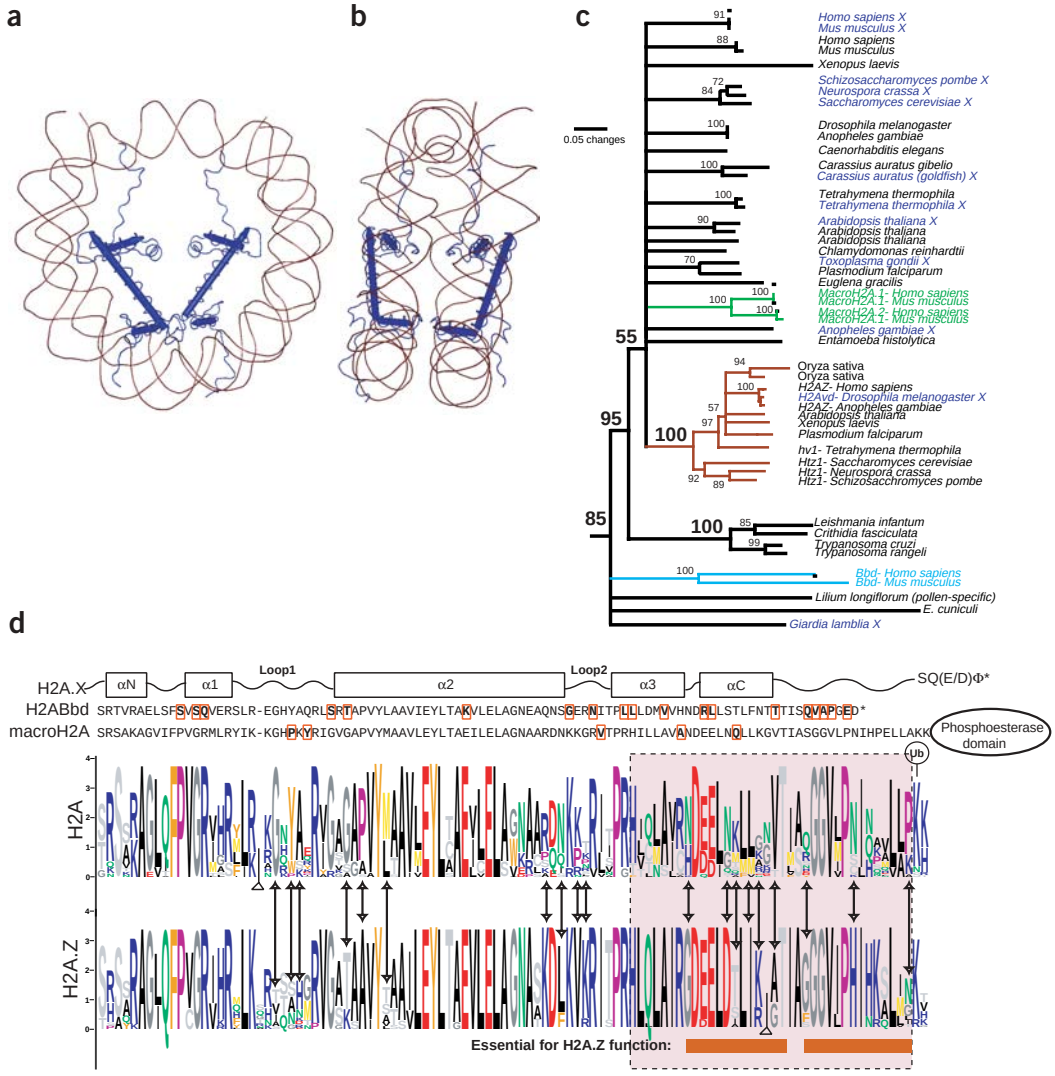

Figure 6 Canonical H2A and variants. (**a**,**b**) Position of H2A within the eukaryotic nucleosome[10]. For clarity, the other three histones are not shown. Note the self-dimerization interface of H2A is in its loop 1 domain. (**c**) Neighbor-joining phylogeny of H2A proteins reveals the ancient split between H2A (black) and H2A.Z (orange lineages)[51], and shows the multiple origins of the H2A.X variant (blue lettering) derived from canonical H2As in all lineages except *D. melanogaster*. MacroH2A is a distinct lineage (green) within the canonical H2As whereas the H2ABbd is a rapidly evolving lineage (teal) that falls outside this grouping. (**d**) An alignment of H2A and H2A.Z variants in Logos format, indicating the three regions where the two differ greatly in their evolutionary constraint (bidirectional arrows). The proposed 'docking domain' is in a dashed box. H2A.X does not differ substantially from H2A except for the addition of a C-terminal motif, SQ(E/D)Φ, where Φ indicates a hydrophobic residue and asterisk refers to C-terminal end of protein. In contrast, macroH2A differs substantially in a few residues from H2A in loop 1 and its docking domain whereas the other vertebrate-specific variant, H2ABbd, has a shorter docking domain and is substantially different from H2A. A C-terminal lysine residue in H2A that is mono-ubiquitinated may play a role in apoptosis.

extended N-terminal and C-terminal tails (**Fig. 6a**,**b**); the N-terminal tail appears to intercalate between the two wraps of the eukaryotic nucleosomal DNA, whereas the C-terminal tail makes extensive contacts with the H3–H4 dimer (referred to as the docking domain) and includes a very short αC helix[10,41]. Several variants have been described[42]. H2A.Z and H2A.X have been found in most eukaryotic lineages, whereas the macroH2A and H2ABbd (for Bar-body deficient) seem to be vertebrate-specific[43,44]. In addition to the H2A variant lineages, the H2A phylogeny also includes outlier lineages in ancestral eukaryotes, in the microsporidian *E. cuniculi* and a lily pollen–specific H2A protein[38].

H2A.X is present in nearly all eukaryotes from *Giardia lamblia* to humans. The defining feature of H2A.X is a C-terminal extension that fits the consensus sequence: SQ(E/D)Φ, where Φ indicates a hydrophobic residue. This motif is crucial for chromatin compaction and repair by nonhomologous end joining (NHEJ). The serine residue in this motif is rapidly phosphorylated in response to, and at the site of, DNA double-strand breaks[45]. These breaks can be induced by ionizing radiation[45], in the course of programmed DNA rearrangements[46,47] or in the initial stages of apoptotic DNA fragmentation[48]. It has been proposed that the phosphorylation of H2A.X and altered chromatin may be a trigger to concentrate repair proteins Rad50, Rad51 and BRCA1 (ref. 49). H2A.X is typically expressed throughout the cell cycle. It may be incorporated at a diluted level in the course of DNA replication throughout the genome, but in response to double-strand breaks, H2A.X may be preferentially deposited at the disrupted nucleosomal arrays. Although these newly deposited H2A.X nucleosomes may trigger the phosphorylation event, it is clear that phosphorylation spreads to a large distance surrounding the original break[50], presumably now affecting the H2A.X nucleosomes that were originally deposited during DNA replication.

It may be critical for the spread of the phosphorylation signal that the nucleosome deposition machinery not be able to discriminate against H2A.X. By being similar in primary sequence to the canonical

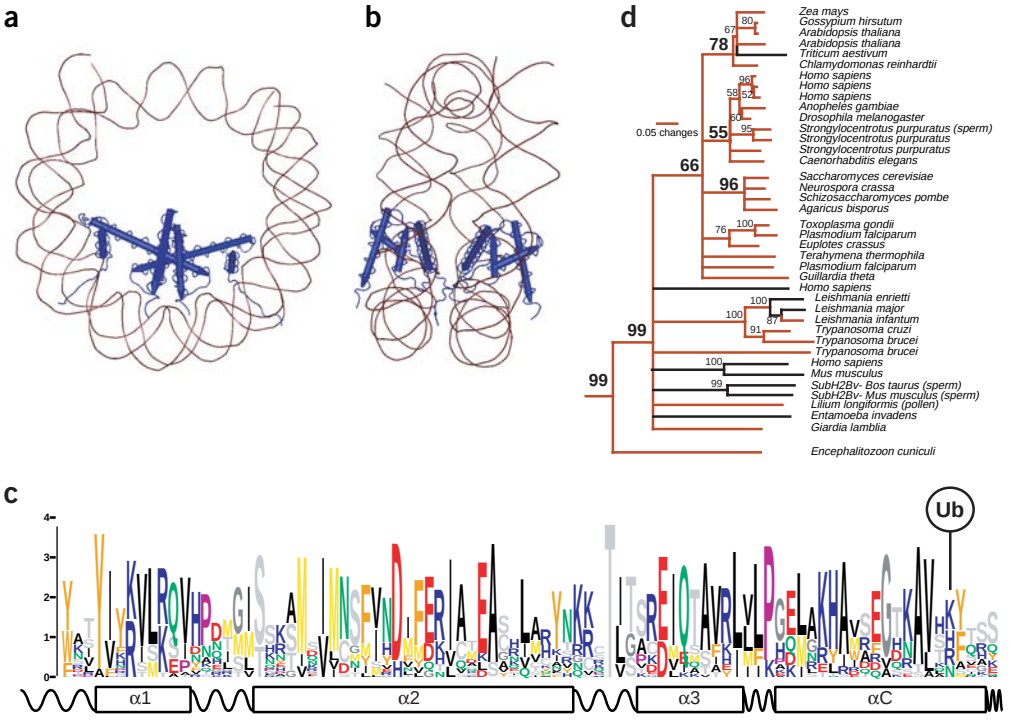

Figure 7 H2B histones. (**a**,**b**) H2B occupies a more peripheral position in the eukaryotic nucleosome with the αC helix on the outer surface[10]. (**c**) An alignment of the various H2Bs shows a more relaxed constraint throughout the HFD. Also indicated is a lysine residue in the C-terminal tail that is mono-ubiquitinated and is responsible for modifications on H3 and H4 proteins. (**d**) A neighbor-joining phylogeny of the various H2Bs shows that this lysine residue is well conserved (orange lineages) except in some orphan lineages, including two in the human genome that are evolving rapidly. One of these orphan lineages, SubH2Bv, is believed to have a role in sperm acrosomes[69].

H2A, H2A.X is packaged along with the canonical H2A. Indeed, it seems from the phylogeny of the H2A proteins (**Fig. 6c**) that the H2A.X variants, defined by the presence of the C-terminal motif, have had multiple evolutionary origins[51]. In every instance, H2A.X is either closely related to (recently derived from) a canonical H2A, or has replaced it entirely, as is the case in fungi and giardia. This leads to an interesting evolutionary question as to how H2A.X is reinvented in the mold of the extant H2A locus. Clearly, this process is not foolproof because the *C. elegans* genome has no H2A.X gene, whereas the *D. melanogaster* genome has atypically selected an H2A.Z gene to be H2A.X as well. This can have profound implications for these two genomes' ability to mediate DNA repair, and perhaps homologous mitotic recombination, which is a byproduct of repair. Thus, whereas *S. cerevisiae* has only H2A.X (no canonical H2A gene) and has high levels of homologous recombination in mitotic cells, *C. elegans* lacks H2A.X, and is poor at this process. Furthermore, mammals and flies have relatively few H2A.X genes and low levels of homologous integration, whereas nearly half of *Tetrahymena thermophila* H2A is H2A.X and it has high levels of homologous recombination[52].

Recently, a non-DNA-repair function of H2A.X has been described. In male mammals, X and Y chromosomes are kept in a transcriptionally silent state in meiotic prophase, cytologically visible as the XY body. H2A.X is implicated in the condensation of the XY body, a step that precedes formation of meiotic recombination–induced double-strand breaks[53].

H2A.Z is highly conserved through most of eukaryotic evolution. Unlike H2A.X, H2A.Z had a single evolutionary origin and has remained distinct from canonical H2A ever since (**Fig. 6c**)[51]. This ancient evolutionary specialization implies that H2A.Z has a role that cannot be substituted by a canonical H2A[54,55]. Studies on H2A.Z have strongly indicated that this variant is critical for maintaining a transcriptionally permissive 'open' state[56] and in protecting euchromatin from encroachment by 'silent' heterochromatin[57]. An interesting variation on the evolutionary origins of H2A.X is observed in *Drosophila* species where an H2A.Z gene (termed H2Avd) has been selected for H2A.X function as well (the C-terminal extension is an unusual SQAY sequence)[58]. The *Anopheles gambiae* genome has a typical H2A.X, so H2Avd must have originated from within the Diptera.

A comparison of H2A.Z to a canonical H2A consensus (**Fig. 6d**) reveals three segments where the two diverge considerably. The largest segment comprises the 'docking domain' near the C terminus, where H2A interacts extensively with the H3–H4 dimer. Functional replacement studies have indicated the essential role that this docking domain has in H2A.Z function[59]. Structural studies have implicated the docking domain in subtly destabilizing the interface between the H2A.Z–H2B and the H3–H4 dimers[41]. This destabilization may be needed to improve accessibility of the nucleosomal DNA to the transcriptional apparatus. 'Looser' packaging of H2A.Z-containing nucleosomes is supported by sedimentation analysis of reconstituted nucleosome arrays[60] and by observations that H2A.Z function is partially redundant with that of nucleosome remodeling factors[61].

In addition to the docking domain, the loop 1 region is quite different between H2A and H2A.Z. Loop 1 is where the two H2A (and H2A.Z) molecules contact each other in the nucleosome[10,11,41]. Structural studies indicate that steric clashes preclude heterodimerization of H2A and H2A.Z in the same nucleosome[41]. In this regard, it is intriguing to note that although canonical H2As are well-constrained in this 'self-interaction' domain, H2A.Z is not especially so. It seems that the only constraint on H2A.Z is to maintain self-interaction, but to avoid heterodimerization with H2A. H2A and H2A.Z also consistently differ in loop 2. This segment might make specific contacts with nucleosomal DNA, or its solvent accessibility might present a unique surface to discriminate between H2A- and H2A.Z-containing nucleosomes. This surface feature might serve to recruit factors involved in chromatin assembly or higher-order packaging[41].

MacroH2A is an unusual vertebrate-specific H2A lineage (**Fig. 6c**)[43]. MacroH2A is enriched in the chromatin of the inactive X chromosome in female mammals (also referred to as the Barr body)[62] and in the transcriptionally silent XY body in male meiosis[63]. X inactivation is the mammalian means by which parity in X chromosomal expression is achieved between males (XY) and females (XX). Two closely related but distinct macroH2A variants have been identified in both human and mouse genomes[64], suggesting that despite their overlapping cytological distribution, some specialization has occurred between macroH2A.1 and macroH2A.2. The macroH2A lineage is also found in birds, which have a different set of sex chromosomes that do not undergo X inactivation[65], suggesting that the macroH2A has only recently adopted a role in X inactivation.

A mysterious aspect of macroH2A evolution is the presence of a 200-residue C-terminal domain, which bears strong homology to proteins found in RNA viruses, eubacteria, archaea and eukaryotic genomes[66]. The yeast homolog of this C-terminal domain (pfam01661) encodes a phosphoesterase that processes Appr-1”-p (an intermediate encountered in the splicing of introns from tRNA) to ADP[67]. The presence of residues that are conserved in the macroH2A domain and other widely dispersed members of this protein family strongly suggests that the domain carries out a closely related enzymatic function. This is the first example of an architectural protein domain covalently linked to a putative modification enzyme.

A recent report provides insight into two roles macroH2A might have in transcriptional silencing[68]. First, the C-terminal domain can interfere with the binding of transcription factors. Second, its H2A domain impedes the activity of nucleosome remodeling complexes. This suggests a tighter packaging of macroH2A-containing chromatin. Based on the precedent of H2A.Z, we examined the amino acid sequence of macroH2A for deviations from the canonical H2A consensus (**Fig. 6d**). Indeed, two residues each in loop 1 and in the docking domain are markedly different. This suggests that H2A and macroH2A do not occupy the same nucleosome, and macroH2A's docking domain could have specialized for even tighter packing.

H2ABbd (Barr body deficient) is the latest H2A variant to be discovered[44]. As its name suggests, the deposition pattern of H2ABbd strongly correlates with trascriptionally active domains and seems to be mutually exclusive to that of macroH2A. Relative to other H2A variants, H2ABbd is evolving rapidly, with branch lengths between human and mouse that are comparable to those separating canonical H2As that diverged about one billion years ago. Whether this is due to relaxed constraints on the H2ABbd protein or adaptive evolution for some novel function remains to be determined. Based on the opposite cytological localization of H2ABbd from macroH2A, H2ABbd-containing nucleosomes are suggested to be 'loose' packing. In this respect, the H2ABbd sequence is quite divergent from both H2A and H2A.Z, particularly so in its shorter docking domain (**Fig. 6d**).

In addition to its histone fold domain, H2B has a C-terminal α-helix (αC) that seems to help mediate its interaction with nucleosomal DNA (**Fig. 7a,b**). The αC helices are on the outer planes of the nucleosomal disc (**Fig. 7b**) and could have a role in stacking of different nucleosomal units into a higher-order structure. H2B is less evolutionarily constrained than H3 or H4 (**Fig. 7c**). As in the H4 lineage, there is limited specialization of H2B proteins. In the phylogeny of H2B proteins, most fungal, plant and metazoan lineages are quite close to each other (**Fig. 7d**), distinct from the orthologous outlier lineages. Two notable exceptions are H2B lineages that seem to have specialized for gametic function. Like its pollen-specific H2A and H3, a divergent pollen-specific H2B protein has been isolated from lily, specialized for the packaging of pollen chromatin[38]. Similarly, a sperm-specific histone, SubH2Bv, is also found in vertebrates (rodent and bovine)[69]. Notably, cytological analyses have shown that SubH2Bv intimately associates, temporally and spatially, with acrosome formation. Despite its similarity to canonical H2B proteins, SubH2Bv is never seen developmentally within the nucleus of the spermatid[69]. It will be interesting to discover what the other partners of this H2B variant are (specifically, whether it is still a heterodimer with H2A) and what role this protein may have in spermiogenesis outside of chromatin packaging. One intriguing possibility is that this version of H2B may represent a storage form of H2B that is important for the 'de-transitioning' from sperm-specific protamines back to canonical histones after fertilization, a role that has been proposed for the H3.3 variant as well.

Histone variants—leaving the 'family' business

Structural investigations of the eukaryotic nucleosome and its putative archaeal ancestor have provided great insights into the evolutionary and biochemical origins of how genetic information is packaged and expressed. Additionally, the identification of reversible and irreversible modifications on N-terminal tails and HFDs have revealed an elaborate network of control that seems to have been established in early eukaryotes. However, these modifications are less likely to alter nucleosomal structure in any fundamental way. There is a growing realization that canonical histones have simply expanded into large multigene families for bulk, indiscriminate packaging of DNA, and may have gotten perhaps too much credit that belongs elsewhere. Their 'orphan' siblings seem to possess most of the discrimination required to permit expression, repair, recombination and even correct segregation of genetic information. With many eukaryotic genomic sequences now available, it is clear that several genomes have chosen to retain variants instead of canonical histones. Following nature's cue, we might refocus and celebrate the variants and even 'deviants'[70] that may have been instrumental for the transition from prokaryotes to eukaryotes.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturestructuralbiology/

1. Sandman, K. *et al.* HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc. Natl. Acad. Sci. USA* **87**, 5788–5791 (1990).
2. Sandman, K., Pereira, S.L. & Reeve, J.N. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell. Mol. Life Sci.* **54**, 1350–1364 (1998).
3. Pereira, S.L. *et al.* Archaeal nucleosomes. *Proc. Natl. Acad. Sci. USA* **94**, 12633–12637 (1997).
4. Musgrave, D.R., Sandman, K.M. & Reeve, J.N. DNA binding by the archaeal histone HMf results in positive supercoiling. *Proc. Natl. Acad. Sci. USA* **88**, 10397–10401 (1991).
5. Sandman, K. *et al.* Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus*. *Proc. Natl. Acad. Sci. USA* **91**, 12624–12628 (1994).
6. Slesarev, A.I. *et al.* Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. *Nucleic Acids Res.* **26**, 427–430 (1998).
7. Ng, W.V. *et al.* Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. USA* **97**, 12176–12181 (2000).
8. Fahrner, R.L. *et al.* An ancestral nuclear protein assembly: crystal structure of the *Methanopyrus kandleri* histone. *Protein Sci.* **10**, 2002–2007 (2001).
9. Decanniere, K. *et al.* Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J. Mol. Biol.* **303**, 35–47 (2000).
10. Luger, K. *et al.* Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
11. Arents, G. *et al.* 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**, 10148–10152 (1991).
12. Ouzounis, C.A. & Kyrpides, N.C. Parallel origins of the nucleosome core and eukaryotic transcription from Archaea. *J. Mol. Evol.* **42**, 234–239 (1996).
13. Eickbush, T.H. & Moudrianakis, E.N. The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry* **17**, 4955–4964 (1978).

14. Fischle, W., Wang, Y. & Allis, C.D. Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* **15**, 172–183 (2003).
15. Wu, R.S., Tsai, S. & Bonner, W.M. Patterns of histone variant synthesis can distinguish G0 from G1 cells. *Cell* **31**, 367–374 (1982).
16. Bosch, A. & Suau, P. Changes in core histone variant composition in differentiating neurons: the roles of differential turnover and synthesis rates. *Eur. J. Cell Biol.* **68**, 220–225 (1995).
17. Thatcher, T.H. *et al.* Independent evolutionary origin of histone H3.3-like variants of animals and *Tetrahymena*. *Nucleic Acids Res.* **22**, 180–186 (1994).
18. Waterborg, J.H. & Robertson, A.J. Common features of analogous replacement histone H3 genes in animals and plants. *J. Mol. Evol.* **43**, 194–206 (1996).
19. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200 (2002).
20. Yu, L. & Gorovsky, M.A. Constitutive expression, not a particular primary sequence, is the important feature of the H3 replacement variant hv2 in *Tetrahymena thermophila*. *Mol. Cell. Biol.* **17**, 6303–6310 (1997).
21. Ausio, J. Histone H1 and evolution of sperm nuclear basic proteins. *J. Biol. Chem.* **274**, 31115–31118 (1999).
22. Hennig, W. Chromosomal proteins in the spermatogenesis of *Drosophila*. *Chromosoma* **111**, 489–494 (2003).
23. Palmer, D.K. *et al.* A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol.* **104**, 805–815 (1987).
24. Palmer, D.K. *et al.* Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl. Acad. Sci. USA* **88**, 3734–3738 (1991).
25. Henikoff, S., Ahmad, K. & Malik, H.S. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
26. Malik, H.S., Vermaak, D. & Henikoff, S. Recurrent evolution of DNA-binding motifs in the *Drosophila* centromeric histone. *Proc. Natl. Acad. Sci. USA* **99**, 1449–1454 (2002).
27. Marzluff, W.F. *et al.* The human and mouse replication-dependent histone genes. *Genomics* **80**, 487–498 (2002).
28. Stoler, S. *et al.* A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* **9**, 573–586 (1995).
29. Buchwitz, B.J. *et al.* A histone-H3-like protein in *C. elegans*. *Nature* **401**, 547–548 (1999).
30. Henikoff, S. *et al.* Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl. Acad. Sci. USA* **97**, 716–721 (2000).
31. Takahashi, K., Chen, E.S. & Yanagida, M. Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science* **288**, 2215–2219 (2000).
32. Talbert, P.B. *et al.* Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* **14**, 1053–1066 (2002).
33. Shelby, R.D., Vafa, O. & Sullivan, K.F. Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. *J. Cell Biol.* **136**, 501–513 (1997).
34. Vermaak, D., Hayden, H.S. & Henikoff, S. Centromere targeting element within the histone fold domain of Cid. *Mol. Cell. Biol.* **22**, 7553–7561 (2002).
35. Malik, H.S. & Henikoff, S. Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics* **157**, 1293–1298 (2001).
36. Chen, Y. *et al.* The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. *Mol. Cell. Biol.* **20**, 7037–7048 (2000).
37. Keeling, P.J., Luker, M.A. & Palmer, J.D. Evidence from β-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol. Biol. Evol.* **17**, 23–31 (2000).
38. Ueda, K. *et al.* Unusual core histones specifically expressed in male gametic cells of *Lilium longiflorum*. *Chromosoma* **108**, 491–500 (2000).
39. Jahn, C.L. *et al.* An unusual histone H3 specific for early macronuclear development in *Euplotes crassus*. *Proc. Natl. Acad. Sci. USA* **94**, 1332–1337 (1997).
40. Douglas, S. *et al.* The highly reduced genome of an enslaved algal nucleus. *Nature* **410**, 1091–1096 (2001).
41. Suto, R.K. *et al.* Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat. Struct. Biol.* **7**, 1121–1124 (2000).
42. West, M.H. & Bonner, W.M. Histone 2A, a heteromorphous family of eight protein species. *Biochemistry* **19**, 3238–3245 (1980).
43. Pehrson, J.R. & Fried, V.A. MacroH2A, a core histone containing a large nonhistone region. *Science* **257**, 1398–1400 (1992).
44. Chadwick, B.P. & Willard, H.F. A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J. Cell Biol.* **152**, 375–384 (2001).
45. Rogakou, E.P. *et al.* DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
46. Chen, H.T. *et al.* Response to RAG-mediated VDJ cleavage by NBS1 and γ-H2AX. *Science* **290**, 1962–1965 (2000).
47. Mahadevaiah, S.K. *et al.* Recombinational DNA double-strand breaks in mice precede synapsis. *Nat. Genet.* **27**, 271–276 (2001).
48. Rogakou, E.P. *et al.* Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139. *J. Biol. Chem.* **275**, 9390–9395 (2000).
49. Celeste, A. *et al.* Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nat. Cell Biol.* **5**, 675–679 (2003).
50. Rogakou, E.P. *et al.* Megabase chromatin domains involved in DNA double-strand breaks *in vivo*. *J. Cell Biol.* **146**, 905–916 (1999).
51. Thatcher, T.H. & Gorovsky, M.A. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. *Nucleic Acids Res.* **22**, 174–179 (1994).
52. Yu, G.L., Hasson, M. & Blackburn, E.H. Circular ribosomal DNA plasmids transform *Tetrahymena thermophila* by homologous recombination with endogenous macronuclear ribosomal DNA. *Proc. Natl. Acad. Sci. USA* **85**, 5151–5155 (1988).
53. Fernandez-Capetillo, O. *et al.* H2AX is required for chromatin remodeling and inactivation of sex chromosomes in male mouse meiosis. *Dev. Cell* **4**, 497–508 (2003).
54. Allis, C.D. *et al.* hv1 is an evolutionarily conserved H2A variant that is preferentially associated with active genes. *J. Biol. Chem.* **261**, 1941–1948 (1986).
55. Jackson, J.D. & Gorovsky, M.A. Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. *Nucleic Acids Res.* **28**, 3811–3816 (2000).
56. Stargell, L.A. *et al.* Temporal and spatial association of histone H2A variant hv1 with transcriptionally competent chromatin during nuclear development in *Tetrahymena thermophila*. *Genes Dev.* **7**, 2641–2651 (1993).
57. Meneghini, M.D., Wu, M. & Madhani, H.D. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**, 725–736 (2003).
58. Madigan, J.P., Chotkowski, H.L. & Glaser, R.L. DNA double-strand break-induced phosphorylation of *Drosophila* histone variant H2Av helps prevent radiation-induced apoptosis. *Nucleic Acids Res.* **30**, 3698–3705 (2002).
59. Clarkson, M.J. *et al.* Regions of variant histone His2AvD required for *Drosophila* development. *Nature* **399**, 694–697 (1999).
60. Fan, J.Y. *et al.* The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nat. Struct. Biol.* **9**, 172–176 (2002).
61. Santisteban, M.S., Kalashnikova, T. & Smith, M.M. Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes. *Cell* **103**, 411–422 (2000).
62. Costanzi, C. & Pehrson, J.R. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393**, 599–601 (1998).
63. Richler, C., Dhara, S.K. & Wahrman, J. Histone macroH2A1.2 is concentrated in the XY compartment of mammalian male meiotic nuclei. *Cyt. Cell. Genet.* **89**, 118–120 (2000).
64. Costanzi, C. & Pehrson, J.R. MACROH2A2, a new member of the MARCOH2A core histone family. *J. Biol. Chem.* **276**, 21776–21784 (2001).
65. Ellegren, H. Dosage compensation: do birds do it as well? *Trends Genet.* **18**, 25–28 (2002).
66. Pehrson, J.R. & Fuji, R.N. Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Res.* **26**, 2837–2842 (1998).
67. Martzen, M.R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
68. Angelov, D. *et al.* The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. *Mol. Cell* **11**, 1033–1041 (2003).
69. Aul, R.B. & Oko, R.J. The major subacrosomal occupant of bull spermatozoa is a novel histone H2B variant associated with the forming acrosome during spermiogenesis. *Dev. Biol.* **242**, 376–387 (2002).
70. Wolffe, A.P. Centromeric chromatin. Histone deviants. *Curr. Biol.* **5**, 452–454 (1995).