

A unified phylogeny-based nomenclature for histone variants

Paul Talbert, Kami Ahmad, Geneviève Almouzni, Juan Ausió, Frederic Berger, Prem Bhalla, William Bonner, Zacheus Cande, Brian Chadwick, Simon Chan, et al.

► **To cite this version:**

Paul Talbert, Kami Ahmad, Geneviève Almouzni, Juan Ausió, Frederic Berger, et al.. A unified phylogeny-based nomenclature for histone variants. *Epigenetics*

Chromatin, *BioMed Central*, 2012, 5 (1), pp.7. <10.1186/1756-8935-5-7>. <inserm-00710902>

HAL Id: inserm-00710902

<http://www.hal.inserm.fr/inserm-00710902>

Submitted on 21 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

REVIEW

Open Access

A unified phylogeny-based nomenclature for histone variants

Paul B Talbert¹, Kami Ahmad², Geneviève Almouzni³, Juan Ausió⁴, Frederic Berger⁵, Prem L Bhalla⁶, William M Bonner⁷, W Zacheus Cande⁸, Brian P Chadwick⁹, Simon W L Chan¹⁰, George A M Cross¹¹, Liwang Cui¹², Stefan I Dimitrov¹³, Detlef Doenecke¹⁴, José M Eirin-López¹⁵, Martin A Gorovsky¹⁶, Sandra B Hake¹⁷, Barbara A Hamkalo¹⁸, Sarah Holec⁵, Steven E Jacobsen¹⁹, Kinga Kamieniarz²⁰, Saadi Khochbin²¹, Andreas G Ladurner²², David Landsman²³, John A Latham¹, Benjamin Loppin²⁴, Harmit S Malik¹, William F Marzluff²⁵, John R Pehrson²⁶, Jan Postberg²⁷, Robert Schneider^{20,28}, Mohan B Singh⁶, M Mitchell Smith²⁹, Eric Thompson³⁰, Maria-Elena Torres-Padilla³¹, David John Tremethick³², Bryan M Turner³³, Jakob Harm Waterborg³⁴, Heike Wollmann⁵, Ramesh Yelagandula⁵, Bing Zhu³⁵ and Steven Henikoff^{1*}

Abstract

Histone variants are non-allelic protein isoforms that play key roles in diversifying chromatin structure. The known number of such variants has greatly increased in recent years, but the lack of naming conventions for them has led to a variety of naming styles, multiple synonyms and misleading homographs that obscure variant relationships and complicate database searches. We propose here a unified nomenclature for variants of all five classes of histones that uses consistent but flexible naming conventions to produce names that are informative and readily searchable. The nomenclature builds on historical usage and incorporates phylogenetic relationships, which are strong predictors of structure and function. A key feature is the consistent use of punctuation to represent phylogenetic divergence, making explicit the relationships among variant subtypes that have previously been implicit or unclear. We recommend that by default new histone variants be named with organism-specific paralog-number suffixes that lack phylogenetic implication, while letter suffixes be reserved for structurally distinct clades of variants. For clarity and searchability, we encourage the use of descriptors that are separate from the phylogeny-based variant name to indicate developmental and other properties of variants that may be independent of structure.

Histones, the basic proteins that wrap DNA into nucleosomes in eukaryotes, are commonly encoded by multi-gene families. Histones fall into five protein families, the core histones H2A, H2B, H3 and H4, and the linker histone family H1. A nucleosome core particle is made by assembling two proteins from each of the core histone families together with DNA. Linker DNA between core particles may be bound by a member of the H1 family. The individual paralogous (non-allelic) genes of a histone family may encode identical proteins, or they may encode related but distinct protein isoforms, commonly referred to as “histone variants”. Though histone variants have been known almost from the beginning of histone

research, we are still discovering the diversity of their roles and functions. Histone variants play critical roles in such diverse processes as transcription, chromosome segregation, DNA repair and recombination, chromatin remodeling, ADP-ribosylation, germline-specific DNA packaging and activation, and even extra-nuclear acrosomal function. Some variants of H2A and H3 have well-studied, specialized functions. Variants of H1 and H2B are common, but much less is known of their functional specialization. H4 variants are few.

The diversity of core histones has led to confusion in naming since their discovery. The current names - H2A, H2B, H3 and H4 - for the ‘canonical’ histones were agreed upon at the Ciba Foundation Symposium in 1975 to simplify and standardize competing names for these proteins based on different methods of extraction [1]. At that time, the first core histone variants of H2A, H2B

* Correspondence: steveh@fhcrc.org

¹Howard Hughes Medical Institute, Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

Full list of author information is available at the end of the article

and H3 had already been described from *Drosophila* [2], from the sea urchin *Parechinus angulosus* [3] and from calf thymus [4,5], respectively. Since then, the numerous variants that have been described from these four basic protein families have been named in a variety of styles, using various combinations of numbers, letters and punctuation. The lack of systematized names can lead to confusion of similar names and incorrect attributions of orthology or common function. For instance, H2Bv from *Plasmodium* can be confused with H2BV from *Trypanosoma*, though the two variants are not closely related. Conversely, the same variant can go by different names in different organisms. A PubMed search for H2A.Z or H2A.Z* gets 269 papers on this variant, but misses 126 that use the names D2, H2Az, Htz1, H2A.F or H2A.F/Z, hv1, H2Av or H2AvD.

At a time when genome sequencing has become routine, the discovery of new variants is common. Indeed, lineage-specific expansions of paralogs confound simple orthology and produce a wealth of unique variants that may or may not warrant specific names to describe them. For example, in humans and mice, there are over 10 different replication-coupled H2A variants and over 10 H2B variants [6], without any clear functional distinctions among them. Faced with this challenge, attendees at the EMBO Workshop on Histone Variants that took place in October, 2011 in Strasbourg, France, found it desirable to develop some consistent rules to apply when naming variants, both to minimize confusion and to aid searching.

Approach and rationale

Here, we begin by surveying the format and rationale of existing histone nomenclature, extracting general usage principles and noting some examples of potential conflicts arising from inconsistent use. We then propose a phylogeny-based nomenclature that utilizes consistent punctuation with letter and number suffixes and (rare) prefixes to arrive at a cogent machine-searchable scheme that is based on expectation of common structure and function through orthology, but which is flexible enough to accommodate new discoveries that will emerge from genome sequencing projects in the coming years.

We limit our discussion to the naming of histone proteins, and not the genes that encode them. Nomenclature rules for genes differ among different organisms. Some gene names may refer to the location and organization of genes or gene clusters that are unique in each species. Furthermore, multiple genes often encode an identical histone variant (for example, the *H3.3A* and *H3.3B* genes of *Drosophila* and humans, and the multiple genes encoding H3.2 and H4 in most animals), leading to ambiguity for a nomenclature basing protein names on gene names or vice versa. We, therefore, leave the

naming of genes to the respective organismal research communities, and accept that histones may have organism-specific protein names based on the genes that encode them, in addition to the names discussed here based only on their amino acid sequence.

Format of existing nomenclature for the core histones

The first core histone variants were denoted with suffixes that included numbers or letters separated by punctuation.

1. Numbers were first used to distinguish paralogs in the same organism, without knowledge of whether or not the proteins were functionally equivalent. Numbers were assigned apparently arbitrarily after chromatographic separation in the case of sea urchin H2B paralogs [3] or according to their family and their mobility in acetic acid/urea/triton X-100 (AUT) gels in the case of mammalian H2a.1, H2a.2, H2b.1, H2b.2, H3.1, H3.2, and H3.3 [7]. More recently, H3 paralogs H3.4 [8] and H3.5 [9] were added, with numbers based on order of discovery rather than electrophoretic mobility.
2. Soon after the introduction of number suffixes, variants followed by letters were introduced to indicate minor subtypes within the H2A family that had previously been overlooked: H2A.X and H2A.Z [10]. Other letter suffixes followed, with or without punctuation: H3t [11]; H3(P) [12]; H2A.Bbd (originally H2A-Bbd) [13,14]; H2BFWT [15]; H2Abd [16]; H3.X, H3.Y [17]; and so on.
3. The letter V was employed in a variety of organisms simply to indicate a difference from another sequence in the same family: H2AvD or H2Av (*Drosophila*) [18,19]; H2Bv (*Plasmodium*) [20]; H2BV, H3V and H4V (*Trypanosoma*) [21,22]. The letter B was used similarly in H3B (*Giardia*) [23].
4. Both numbers and letters have been used together as suffixes when recognized variants have paralogs: H2AL1 and H2AL2 (mouse) [24]; H2A.Z-1 and H2A.Z-2 (vertebrates) [25]; H2Abd1_c and H2Abd2_a/b (rotifers) [16]; H3v1 to H3v10 (*Stylonychia lemnae*) [26,27]; H2Asq.1 to H2Asq.3 (*Oikopleura dioica*) [28].
5. Both numbers and letters have been used for splice variants: macroH2A1.1 and macroH2A1.2 (vertebrates) [29]; H2A.Za to H2A.Zc (*Oikopleura dioica*) [28].

Some histone variants have been designated with a prefix.

1. Prefixes have marked variants that are divergent in sequence and developmentally restricted: TH2B [30]

- or hTSH2B [31]; gH3 [32], gC3 [33], and leH3 [34], and so on. They have also designated variants that are structurally and/or functionally distinct: macroH2A or mH2A [35]; CenH3 [36].
2. Some variants have been assigned both a prefix and a suffix: SubH2Bv [37], soH3-1 and soH3-2 [34], and so on.
 3. Specifying the histone variant of an individual species has often been accomplished with a prefix: *HsH2B*, *OdH4*, *PfCenH3*.
 4. Often descriptors have preceded variant names, such as for variants specific to certain developmental or cell-cycle stages: for example, early H2A, cleavage stage (CS) H2B [38], replication-coupled (RC) H3 and generative cell (GC) H3s [34].

Overlaid on top of these designations has been imposed the Brno nomenclature for indicating modifications [39], which involves suffixes designating the kind and position of the modified residue and the nature of the modification, all without punctuation, for example, H3S10ph; H3K27me3; H2A.Z2K4acK11ac.

It is evident from these examples that there has been little consistency in the use of punctuation, capitalization, or prefix versus suffix to designate histone variants. There has been some consistency in the use of numbers for paralogs that have not been functionally differentiated. Some names that bear an uncomfortable resemblance to each other (for example, H2A.Bbd and H2AAbd, H2Bv and H2BV) have been assigned to completely distinct variants.

Phylogeny and historical usage

Our goal is to make guidelines that would result in logical, simple, consistent, searchable and informative names that preserve as much of the historical usage as possible. Our approach to such guidelines begins by favoring the use of information obtained from histone phylogenies to guide the creation of a logical naming system for these proteins or, at least, to be consistent with such a naming system. Indeed, the reconstruction of phylogenies has proven to be an excellent strategy in predicting structural and functional features in histones. This argument is illustrated with H2A.Z, where the detailed biochemical analysis of vertebrate-specific H2A.Z-1 and H2A.Z-2 fractions [40] was driven by previous phylogenetic analyses that pointed to their functional differentiation [25]. Therefore, a naming system based on the phylogenetic relationships among histone members within different families can help place variants in a structural, functional and evolutionary framework.

Although histones have been classically considered as archetypal examples of slow-evolving proteins, some histone variants do evolve very quickly, particularly those specifically associated with the germinal lineage (that is,

H2A.Bbd, sub-acrosomal H2Bs, and so on). The presence of such heterogeneity in evolutionary rates mirrors heterogeneous selective constraints in diverse functional backgrounds, operating over the genetic diversity generated through birth-and-death during the long-term evolution of histones [25,41]. Genome sequences already tell us that there are many amplifications of variants that are particular to specific lineages [6,28,38,42,43]. Consequently, we recognize that the phylogeny and orthology of histone variants are not always clear. In a family of proteins encoded by multicopy genes, names will, therefore, commonly be used to specify a class of related orthologs and paralogs rather than a specific protein sequence (for example, H2A.Z, not *OdH2A.Zb*), and flexibility to accommodate phylogenetic uncertainty is necessary.

We also recognize that changing the names of proteins is inherently confusing and disruptive to literature searches. While some name changes are necessary to create a coherent system to guide the naming of new variants, renaming should be minimized. In some cases, compromises in naming conventions to accommodate historical usage are preferable to complete logical consistency. These considerations prevent a strictly phylogenetic approach, but encourage a flexible approach that aims to incorporate phylogeny where practical.

The guidelines we propose are described below and summarized in Table 1.

Capitalization

From the point of view of effectively designating histone variants, we see no reason to prefer upper or lower case, or to differentiate between them. Most search engines do not distinguish upper and lower case. However, there are two reasons to prefer upper case as the default in naming variants. First, some genetic nomenclature systems differentiate proteins from genes by using uppercase for proteins. Second, the Brno nomenclature for histone modifications specifies that the protein and its modified residue(s) are upper case, while the modifications are lower case. The use of uppercase in H2A, H2B, H3, H4 and their suffixes would aid implementation of the Brno nomenclature without affecting the ability to search.

Traditionally, some histone variant prefixes (for example, macroH2A) have not been capitalized, while others have been (for example, TH2B). The use of lower case in prefixes seems unlikely to create confusion with Brno modification nomenclature or gene names, and makes the Ciba core histone family designation more visually prominent. We, therefore, prefer the use of lower case for prefixes (for example, macroH2A, subH2B, cenH3). The use of lower case should not affect searching for existing literature that uses uppercase prefixes.

Table 1 Summary of nomenclature guidelines

Naming feature	Recommendation	Examples
Core histone name	Use in an inclusive sense for the protein family. Specify subgroups with a descriptor, prefix, letter suffix, or number suffix.	'H2A can be ubiquitylated.' 'H3 can be methylated on K4.'
Capitalization	Upper and lower case are equivalent in meaning, but upper case is preferred for designating core histones, their suffixes, and modifiable amino acids. Use lowercase for modifications and for prefixes.	H3.3K4me3, H2BK123ub1, cenH3
Descriptors	Descriptors can be used before the core histone name to specify a feature, group variants developmentally or functionally, indicate the species of origin, or other uses. There should be a space between the descriptor and the core histone name. There is no requirement that a descriptor specifies a clade.	RC H2A, early H4, testis-specific H3.4 or TS H3.4, <i>Hs</i> H2A. X or human H2A.X, GC H2As, oocyte H1s
Prefixes	These should be few in number and specify a structurally distinct clade of a core histone that is universal or characteristic of a high-level taxonomic clade. Lower case is preferred for prefixes.	macroH2A, cenH3, subH2B
Letter suffixes	These should be preceded by a period (.) and specify a structurally distinct monophyletic clade of a histone family (exception: H2A.X). A suffix may be applied judiciously at any taxon level.	H2A.Z, H3.X, H2A.B
Number suffixes	These should be preceded by a period (.) and specify a particular variant of a core histone, without constraint as to distinctiveness and without implication as to phylogeny. Number suffixes should be assumed to be species-specific, but it is convenient to name variants in related species consistently where unique orthologies are clear. A number suffix should be the default designation of new variants.	H3.5, H2A.1, macroH2A.2, H1.0
Punctuation	Use a period (.) after core histone names to indicate a subtype (letter or number suffix). Use additional periods as necessary to separate finer divisions of subtypes. A period is equivalent to a branch point in a phylogenetic tree.	H2A.Z.1, H2A.L1
Splice variants	Use a period (.) before a splice variant number. Treat the same as paralog number suffixes, except that a lowercase 's' may precede the number to indicate that the isoform is a splice variant.	macroH2A.1.2, H2A.Z.s3
Synonyms	For names changed by this nomenclature, refer to both old and new synonyms in the abstract of papers to facilitate literature searches. Optional descriptors can aid identification.	'Avian H1.0, also known as H5'

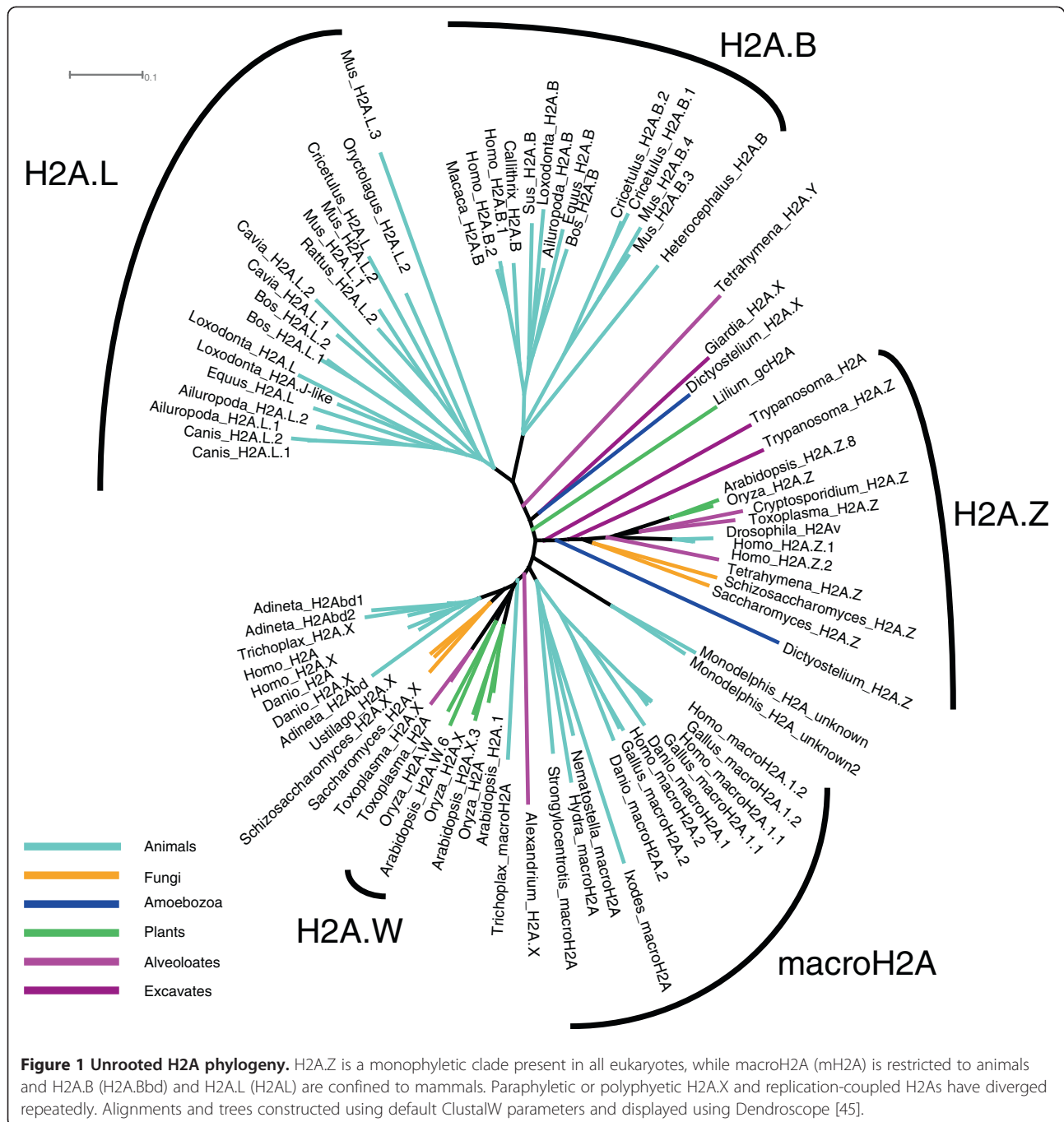
Prefixes

The majority of histone variants have been designated by suffixes, and only a handful have used prefixes. Of these prefixes, only macroH2A, cenH3, and TH2B have an extensive literature.

'macroH2A' or 'mH2A' describes a well-delimited clade of proteins containing the 'macro-domain' (Figure 1), which sets these proteins apart from other H2A variants [35]. macroH2A is found in diverse animal phyla, including such basal lineages as cnidarians and placozoans [44], with subtypes and splice variants in vertebrates [29]. It is likely to have been ancestrally present in animals, but has been lost in several lineages, including *Drosophila* and *Caenorhabditis*.

'CenH3' was proposed as a functional (rather than structural) class [36] because the monophyly of centromeric H3 proteins is uncertain [27,46,47], although monophyly seems the most parsimonious hypothesis. Either way, cenH3s form a distinct group with recognizable structural features [47,48].

The variant SubH2Bv was described in bull sperm, and potential orthologs with characteristic divergent histone fold domains have been identified in mice (H2BL1) and other mammals [24,37]. This apparently rapidly evolving protein variant (Figure 2) is reportedly found in the subacrosome of primates, rodents, and marsupials as well as bovids [37,49]. The use of the prefix 'sub' for this family seems appropriate, since residence in the subacrosome



rather than the nucleus is a distinctive property that appears to be distributed throughout mammals. We recommend dropping the superfluous suffix 'v'. Specific recommendations for renaming variants are summarized in Table 2.

From these examples, we suggest that prefixes be used sparingly to designate structurally distinct families of variants with wide distribution within a 'high-level' clade. 'High-level' may be taken to correspond to traditional

classes, phyla, kingdoms or more inclusive clades. Of existing variants not designated with prefixes, H2A.Z would be an obvious candidate for prefix designation by our criteria if it were not the prototypical example of using letter suffixes for variants.

'TH2B' describes a mammalian testis-specific H2B, which can dimerize with H2A1L1/L2, and may form sub-nucleosomal particles in condensed spermatids [24,30]. hTSH2B is the human ortholog of TH2B in rodents [31]

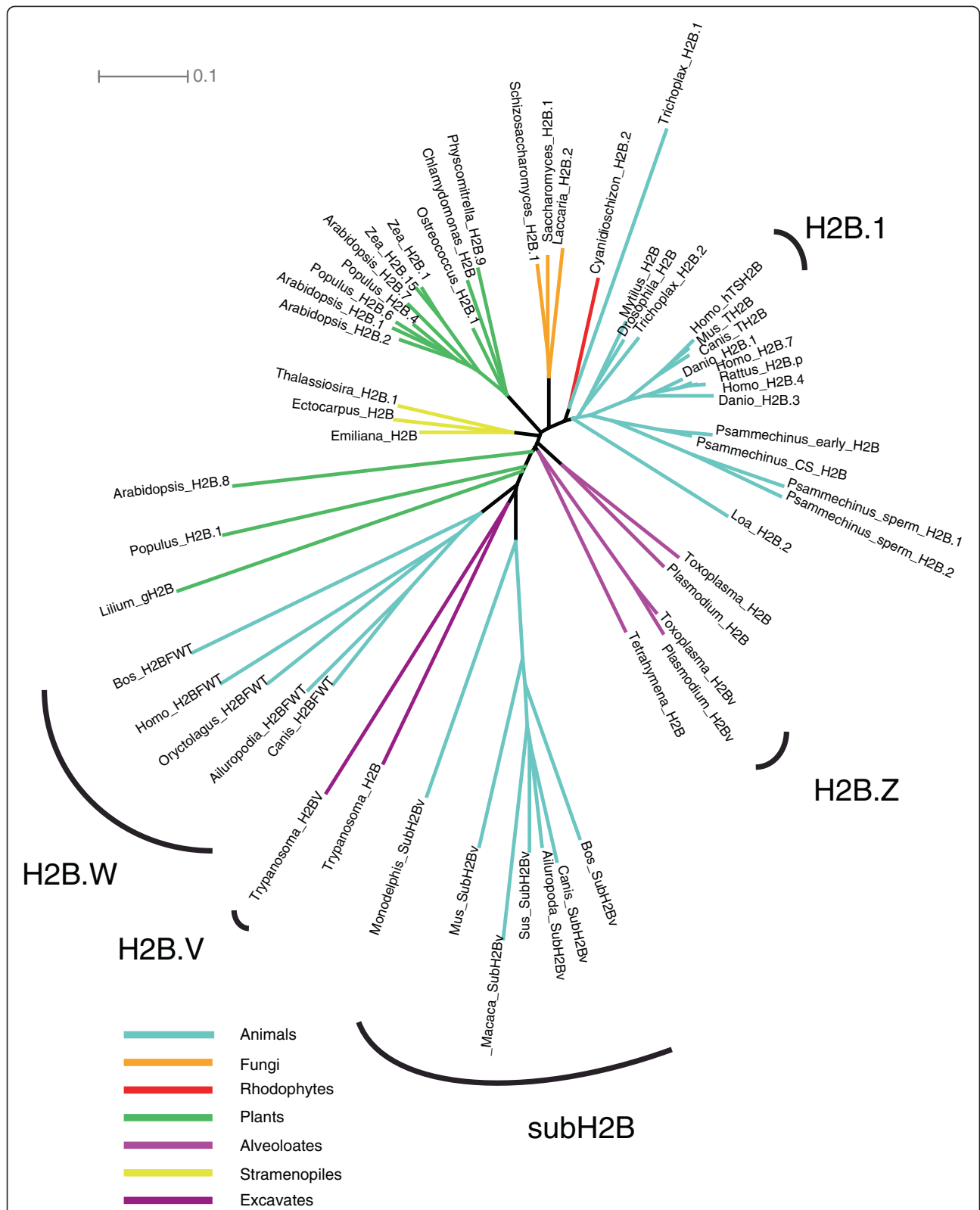


Figure 2 Unrooted H2B phylogeny. TS H2B.1 (TH2B), H2B.W (H2BFWT) and subH2B (SubH2Bv) are mammal-specific clades. Highly divergent generative cell H2Bs in plants do not form a clear clade. Apicomplexan H2Bv does not appear to be related to trypanosome H2Bv, despite the fact that both are thought to interact with H2A.Z.

Table 2 Specific name change suggestions for histones

Old name	Organism(s)	New unified name
H2A (with SPKK motifs)	plants	H2A.W
H2A.Bbd	mammals	H2A.B
H2Abd1, H2Abd2, H2Abd	bdelloid rotifers	(bdelloid) H2A.1 to (bdelloid) H2A.3
H2AL	mammals	H2A.L
H2Av, H2AvD, D2, hv1, Htz1p	<i>Drosophila</i> , <i>Tetrahymena</i> , <i>Saccharomyces</i>	H2A.Z
SubH2Bv	mammals	subH2B
H2BL1	mammals	subH2B
H2Bv	apicomplexans	H2B.Z
H2BV	trypanosomes	H2B.V
H2BFWT	mammals	H2B.W
TH2B, hTSH2B	mammals	(TS) H2B.1
H3(P)	<i>Moneuplotes</i>	H3.P
H3t	mammals	(TS) H3.4
H3v1 to H3v10	<i>Stylonychia</i>	H3.1 to H3.10
H3V	trypanosomes	H3.V
H3.X	human	H3.Y.2
H3.Y	human	H3.Y.1
H4V	trypanosomes	H4.V
H1°	animals	H1.0
H5	birds	H1.0
H1δ	echinoderms	H1.0
H1t	mammals	(TS) H1.6
H1T2	mammals	(TS) H1.7
H1oo	mammals	(OO) H1.8
H1s1	mammals	(TS) H1.9
H1x	vertebrates	H1.10
B4	frogs	(Amphibian) H1.4

Parentheses () indicate an optional descriptor.

and should be designated by the same name. Although TH2B is somewhat diverged from other vertebrate H2Bs, it falls well within the clade of animal replication-coupled H2Bs (Figure 2), and is primarily distinguished by its testis-specific expression. It is the first H2B gene encoded in the major mammalian histone gene cluster, and is designated as ‘type 1’. We, therefore, suggest that mammalian testis-specific H2B.1 (TS H2B.1) would be a better designation for this variant, where ‘H2B.1’ is the name and “mammalian testis-specific” is a descriptive phrase that helps alert readers to its properties.

Other existing prefixes include those used for the five H3 variants expressed in the generative cell (GC) of *Lilium* [34]. Of these, gH3 and leH3 have deletions in the histone fold domain and are probably non-functional, although they

might have non-nucleosomal functions like SubH2Bv. gH3 makes a chromatin protein, but there is no evidence at present that it is not a *Lilium*-specific variant (Figure 3A). soH3-1 and soH3-2 are subtypes of H3.3. All five of these variants should be designated by paralog numbers until such time as they are demonstrated to represent widespread subtypes.

Descriptors

The continued use of the descriptive phrase ‘GC H3 variants’ for these *Lilium* H3s [34] and for similar variants in *Arabidopsis* and other plants would be consistent with our proposal. In this case, ‘GC’ is not a prefix, but is separated by a space from the name, and acts as a descriptor to specify that the variants are found in generative cells. Such descriptors may be applied to describe functional, stage-specific or other groupings of variants that do not necessarily form phylogenetic clades. The (roughly) corresponding descriptor for animals is testis-specific (TS). TS variants appear to be widespread, rapidly evolving and polyphyletic in animals [9,28,38]. The same seems likely to be true of GC variants in plants. We encourage the use of flexible, detached descriptors rather than permanent prefixes or suffixes to specify developmental stages or cell-type specific expression of histone variants, because subsequent work may reveal a wider developmental deployment for a variant than was initially discovered. We specifically encourage the use of descriptors, such as TS, to specify testis-specific histones, including TS H2B.1 in mammals and the non-orthologous TS H2Bs found in other groups of animals [28,38].

Descriptors may also be used to designate the organism. As with other descriptors, a space should be maintained before the variant name to facilitate machine searches and reduce possible confusion (*Hs* H2A.Z or human H2A.Z, not HsH2A.Z). Under our scheme, the three classes of bdelloid rotifer H2Abd, which all have long but apparently unrelated tails, might be better designated with a species or group descriptor and paralog numbers (*Adineta vaga* H2A.1 to H2A.3 or bdelloid H2A.1 to H2A.3) since it is presently unclear whether these variants form a clade or are polyphyletic (Figure 1). If it can be shown that they form a clade, then a common prefix or letter suffix would be appropriate.

While descriptors are intended to be flexible and unregulated, we suggest that a practical order for the use of multiple descriptors might be (organism or group) (developmental stage and/or tissue) (other) before the Ciba designation and any suffix. Descriptors are optional, but are encouraged where they are informative, especially upon first occurrence of a variant name in a manuscript.

Suffixes

Letter suffixes function essentially the same as prefixes in designating structurally distinct variants, except that

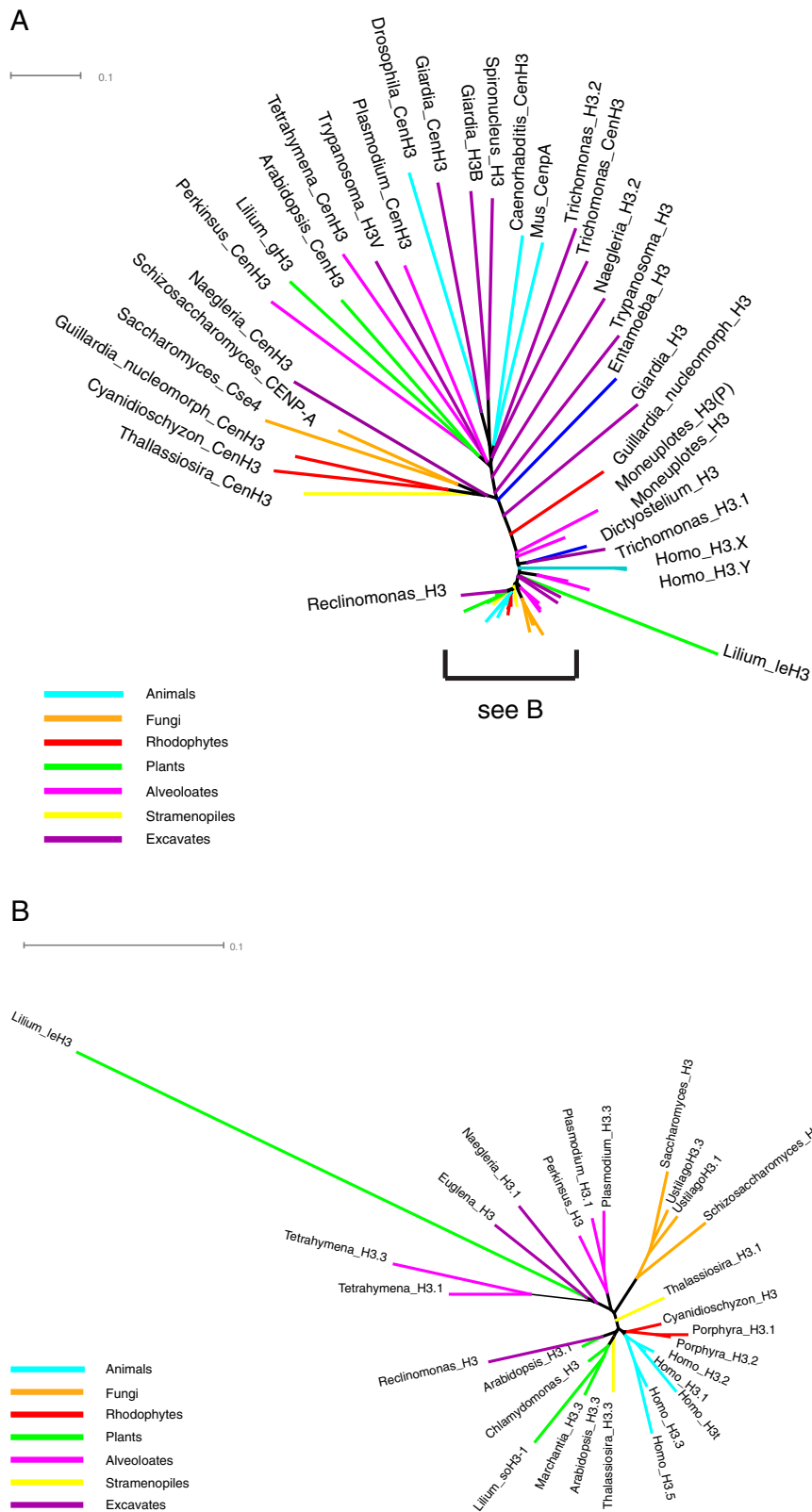


Figure 3 Unrooted H3 phylogeny. (A) cenH3s are not clearly separable from the divergent H3s of excavates and of plant generative cells. **(B)** Replication-independent H3.3s and replication-coupled H3s have diverged repeatedly in different lineages.

they need not characterize a high-level clade. We prefer the use of single letter suffixes preceded by a period, as in the prototypes H2A.X and H2A.Z. Thus H2A.W is proposed for the plant-specific clade of H2As (Figure 1) with putative minor-groove-binding motifs (SPKK) in their tails that wrap more DNA than other H2As [50]. We further suggest H3.P for *Moneuplotis* H3(P), H2A.B for mammalian H2A.Bbd, and H2B.W for mammalian H2BFWT. The continued use of 'H2A.Bbd' would still be searchable with 'H2A.B*', so in this case the traditional name could still be used by those who prefer it, though we suggest transitioning to 'H2A.B'. 'H2BFWT' has only two PubMed hits [15,51], and NCBI already refers to this family as 'member W' or 'type W-T', so we think it would be minimally disruptive to change its designation to 'H2B.W'. Although it may eventually be necessary to use multiple-letter suffixes, we prefer to stick with single letters until they become insufficient.

H3.X and H3.Y are part of a primate-specific clade that is distinct from other H3s, but they are highly similar to each other; so far, the *H3.X* gene has not been shown to express protein [17]. We suggest designating these variants H3.Y.1 and H3.Y.2 to better reflect their close relationship, and to save 'H3.X' for a verified protein with a distinct structure. In general, in order to avoid running out of unique letter suffixes, we suggest that the default for designating new variants should be the assignment of paralog numbers, with the assignment of letters judiciously applied to variants with obviously unique features. For example, the H2A.Y variant from *Tetrahymena* clearly warrants a letter suffix because of its unique leucine-rich-repeat domain and phosphatase regulating activity [52]. In contrast, two human H2A.Bbd variants differ by only a single amino acid of unknown significance and should be distinguished through paralog numbers: H2A.B.1 and H2A.B.2.

We advocate the continued use of paralog numbers preceded by a period to designate individual variants of one family without implication as to their phylogenetic interrelationships or functions. Paralog numbers should be assumed to be organism-specific unless otherwise indicated. Thus H2A.1 of humans would not be assumed to be the same as H2A.1 of *Arabidopsis* or of *Oikopleura*. However, the assignment of paralog numbers in one organism should be consistent with paralog numbers in related organisms if unique orthologies are clear. For example, vertebrates have two macroH2A proteins, one designated macroH2A1 or macroH2A.1 that is orthologous to the macroH2A.1 originally described from rat liver [35], and one designated macroH2A2 that is orthologous to the macroH2A2 first described from humans [53]. There is no need to assign paralog numbers sequentially if there is some phylogenetic, mnemonic or other reason to prefer a non-sequential number, such as correspondence to the

names of orthologs, pre-existing gene numbers, or to names assigned by gene organization.

The (non-centromeric) H3 variants constitute a special case because they are few in number and have the possibility of a fairly complete phylogeny [27,54-57], and also because of the historical usage of the human paralog numbers H3.1, H3.2 and H3.3. Phylogenetic analysis has inferred that both replication-coupled (RC) H3s and replacement or replication-independent (RI) H3s have most likely arisen polyphyletically (Figure 3B). Current usage often distinguishes between 'H3' for RC forms and 'H3.3' for replacement forms in a variety of organisms [7,58-62]. In contrast, ascomycetes generally have only a single form which is H3.3-like but is usually referred to simply as 'H3' [63]. Unqualified 'H3' is also used in an inclusive sense for all H3 variants in the many contexts in which the variants cannot be distinguished. We encourage the use of this inclusive meaning for 'H3' (without descriptors), which would, therefore, continue to apply to ascomycetes, and the use of an organism-appropriate paralog name (H3.1, H3.2, and so on.) or descriptors such as 'RC' to indicate replication-coupled H3 variants. H3.1 is a more recent mammal-specific divergence from H3.2, which is the RC H3 variant found throughout animals [27].

Following the usage in animals, 'H3.3' has been applied in plants [42,58] and alveolates [57,60,61] to indicate RI variants. Given the likelihood of independent divergences, these variants are neither more nor less orthologous to animal H3.3 than their RC counterparts, but it would be highly inconvenient to alter this practice. This highlights a useful feature of paralog numbers: since paralog numbers are not intended to imply unique corresponding orthology across organismal groups, the use of 'H3.3' in multiple kingdoms does not misrepresent orthology, but functions as a well-established way of indicating RI variants in a variety of organisms that is shorter than using a descriptor, such as 'replacement' or 'RI' H3. RI and RC variants within an organism typically differ in residue 31 (and whether it can be phosphorylated) as well as residues 86 and 89, as shown in Table 3, but distinguishing residues vary in different organisms and caution is advised in designating 'H3.3' in less well-studied eukaryotic kingdoms.

Punctuation

Punctuation and no punctuation are both currently used when appending sub-designations to basic histone families. Punctuation is convenient for separating numeric paralog designations from the alphanumeric Ciba names of the histone families (for example, H3.3, not H33). Periods, dashes, and parentheses have been used without any distinction in meaning, but the period is the most common form of punctuation in histone variant names, and is

Table 3 Sequence variation in 'H3.3' and 'H2A.X' variants

Kingdom	Organism	Histone Variant	Residue 31 -	Residues (85)86-89
Animals	<i>Homo</i>	H3.3	S	AAIG
	<i>Homo</i>	H3.2	<u>A</u>	<u>SAVM</u>
Fungi	<i>Saccharomyces</i>	H3	S	SAIG
Plants	<i>Arabidopsis</i>	H3.3	T	HAVL
	<i>Arabidopsis</i>	H3.1	<u>A</u>	<u>SAVA</u>
Rhodophytes	<i>Porphyra</i>	H3.3?	S-	TAVL
	<i>Porphyra</i>	H3.1?	<u>VG</u>	<u>SAVL</u>
Alveolates	<i>Tetrahymena</i>	H3.3	VS	QAIL
	<i>Tetrahymena</i>	H3.1	<u>AT</u>	<u>SAVL</u>
Heterokonts	<i>Thalassiosira</i>	H3.3?	TA	STAVL
	<i>Thalassiosira</i>	H3.1?	<u>AT</u>	<u>GSAVL</u>
Amoebozoa	<i>Dictyostelium</i>	H3.3?	STQP	AAIQ
	<i>Dictyostelium</i>	H3.1?	<u>VNEV</u>	<u>AAIE</u>
Excavates	<i>Euglena</i>	H3	A	NAIL
Residues 109-				
Animals	<i>Homo</i>	H2A.X	PNIQAVLLPKKSATVGPKAPSGGKKATQ ASQEY	
	<i>Homo</i>	H2A.2.2	PNIQAVLLPKKTSHKPGKNK	
Fungi	<i>Saccharomyces</i>	H2A.X	PNIHQNLLPKKSAKATK ASQEL	
Plants	<i>Arabidopsis</i>	H2A.X.3	PNIHQTLPSKVGKNKGDIGS ASQEF	
	<i>Arabidopsis</i>	H2A.1	PNIHNLPLPKKAGASKPQED	
Rhodophytes	<i>Griffithsia</i>	H2A.X	PNIHQVLMPRKKTGDA ASQEV	
	<i>Cyanidioschyzon</i>	H2A	PNIHAVLLPKKAKAGE	
Alveolates	<i>Tetrahymena</i>	H2A.X	PNINPMLLPSKSKKTESRGG ASQDL	
	<i>Tetrahymena</i>	H2A.1	PNINPMLLPSKTKKSTEPEH	
Heterokonts	<i>Phaeodactylum</i>	H2A.X	PNIHAILLPKKTIKTKGP SQDY	
	<i>Phaeodactylum</i>	H2A.3	PNIHAILLPKKSQPTK	
Amoebozoa	<i>Dictyostelium</i>	H2A.X?	PTPQQSTGEKKKPSKAAEGSS SQIY	
	<i>Dictyostelium</i>	H2A	PTPQSNTGKSKKATSKKS	
Excavates	<i>Giardia</i>	H2A.X	RSAKEGREGKGS HRSQDL	
	<i>Trypanosoma</i>	H2A	PSLNKALAKKQKSGKHAKATPSV	

RI 'H3.3' variants and 'H2A.X' variants are paraphyletic or polyphyletic, but have recognizable sequence features. Upper panel: Divergence between RI and RC H3 variants usually involves differences at residue 31 and residues 86 to 89. Lower panel: 'H2A.X' variants are distinguished from related H2A variants by bearing a consensus SQ(E/D)Φ phosphorylation motif at the C-terminus. Residue numbers refer to the human H3.3 and H2A.X protein sequences, and to orthologous positions in other variants.

essential for finding the relevant literature on these variants in PubMed. In addition, the use of other special characters, such as parentheses, dashes, slashes, superscripts and subscripts, and so on, can complicate searches. We suggest the use of a period to separate each appended sub-designation (except unpunctuated modification designations, in keeping with the Brno nomenclature).

In both letter and number suffixes, the period functions essentially to designate a branchpoint in a phylogenetic tree: H2A.Z, H2A.Y and H2A.1 represent different branches of

the H2A family. The use of the original name 'macroH2A.1' for vertebrate macroH2A1 [35] would extend this principle to the vertebrate branches of the macroH2A family, if the corresponding form 'macroH2A.2' were used for the original form 'macroH2A2' [53].

This logic of designating branchpoints with a period can be extended to subsequent branchpoints as needed. Thus, vertebrate H2A.Z-1 and H2A.Z-2 can be represented as H2A.Z.1 and H2A.Z.2, indicating the two branches of the H2A.Z subfamily. Similarly, mouse

H2AL1 and H2AL2 can be designated H2A.L.1 and H2A.L.2, and *Oikopleura* H2Asq.1 to H2Asq.3 can be H2A.Q.1 to H2A.Q.3 or similar designation.

Although paralog number suffixes are not generally intended to mark clades, in some cases variants that are not distinctive enough to warrant a letter suffix, nevertheless fall into recognizable clades with subtypes that can be described using a period and additional suffix. For example, human H2A.1 and H2A.2, as originally defined electrophoretically, actually represent two subfamilies of H2A variants that differ by whether they have leucine or methionine at position 51. By designating the individual variants in these subfamilies using an additional branchpoint (for example, H2A.1.6 or H2A.2.3), individual variants can be uniquely designated while retaining H2A.1 and H2A.2 for the original subfamilies as defined electrophoretically.

The same logic could be applied to any case when an organism has multiple similar variants that group into subfamilies, but in some cases such phylogenetic detail may be more distracting than informative. Mammalian H3.1 is clearly derived from animal H3.2, but there seems little advantage in designating mammalian H3.1 and H3.2 as 'H3.2.1' and 'H3.2.2', although both pairs of designations would be allowable under our nomenclature guidelines. Similarly, where there is more than one H3.3-like variant in an organism, as is the case for *Caenorhabditis* and many plants [64], it will usually be simplest to assign different paralog numbers to the individual variants, presumably with 'H3.3' assigned to the most abundant or appropriate such variant. For example, in *Arabidopsis*, the germline-specific RI variant known as HTR10 (a gene-derived name) or AtMGH3 (a prefix of the type we discourage here) [42,65] might be designated 'H3.10' to distinguish it from the ubiquitous RI variant 'H3.3', while avoiding the equally correct but more cumbersome designation 'H3.3.10'. In general, we believe that multiple numeric suffixes can become confusing and that shorter names are preferable, unless there is a compelling reason to provide a name that incorporates detailed phylogeny. In names, as in phylogenetic trees, clarity is usually more important than representing every known branchpoint.

We see no reason to treat splice variants differently than paralogs, so the same branchpoint logic can be usefully applied to splice variants: macroH2A1.1 and macroH2A1.2 can be designated as macroH2A.1.1 and macroH2A.1.2, while *Oikopleura* H2A.Za to H2A.Zc [28] would become H2A.Z.1 to H2A.Z.3. While the latter might lead to confusion with vertebrate H2A.Z.1 and H2A.Z.2, this ambiguity is inherent in the use of organism-specific paralog numbers, which seems unavoidable given the ubiquity of lineage-specific expansions of variants. When it is desirable to distinguish splice variants

from paralogs for clarity, we suggest allowing the use of the lowercase letter 's' (for 'splice variant') before the splice variant number, for example macroH2A.1.s1 or H2A.Z.s2.

The imposition of this formal punctuation is intended to apply to written designations, not to impose a stilted formality to speech. In common usage, (.) is pronounced "point" before numbers (for example, " $\pi = 3.14$ " or "histone H3.3") and "dot" before letters (for example, "NIH.gov"). The "dot" is often dropped in pronouncing variants like H2A.Z or H2A.X. We have no intention of interfering with these or other patterns of natural speech. We only seek consistent punctuation in written names to achieve uniform spelling rules that aid searching and express phylogenetic relationships.

Synonyms and homographs

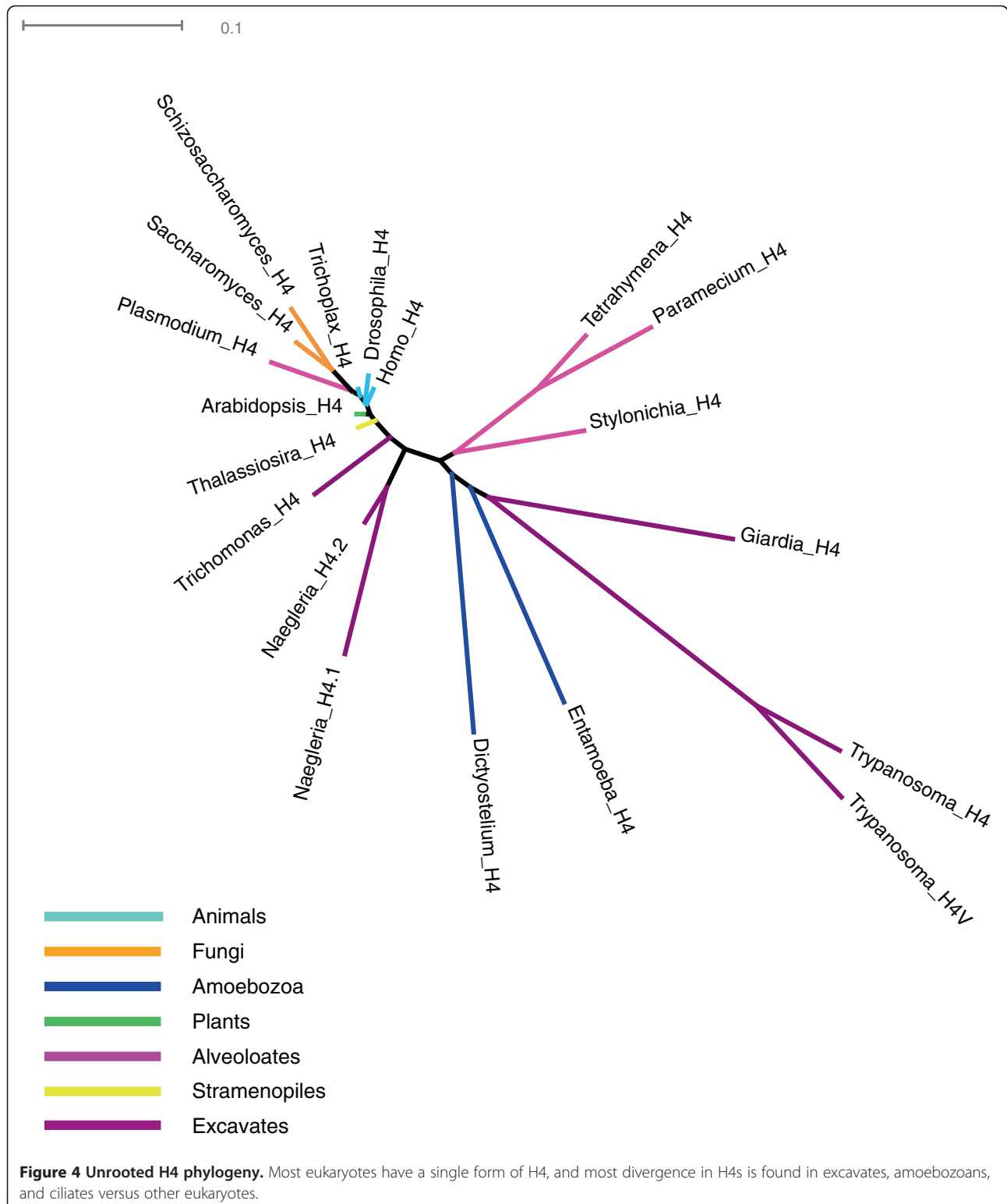
The use of alternative names for the same or the orthologous variant should be discouraged, except to list synonyms. Thus, we favor using H2A.Z in preference to *Saccharomyces* Htz1p, *Tetrahymena* hv1, or *Drosophila* H2Av/H2AvD/D2. H2Av has a convergent phosphorylation motif that allows this protein to function similarly to H2A.X. This does not alter the fact that this protein is a legitimate H2A.Z, but it suggests that the alternative name H2A.Z.X might be useful for discussing this protein in contexts that concern its phosphorylation. The phosphorylated state can either be denoted according to the Brno nomenclature (H2A.ZS138ph) or as γ H2A.Z.X, in parallel with the usual γ H2A.X.

The cases of CENP-A [66], Cse4p [67] and some other centromeric H3s are somewhat exceptional in that these names are earlier and have priority over cenH3, and are well-established in animal and yeast literature. The established use of multiple names was part of the rationale for creating a functional category to apply to all centromeric H3s regardless of monophyly. PubMed treats CENP-A and cenH3 as synonyms, and Cse4p is well-known to centromere researchers, so the (im)practical consequences of synonymy are largely ameliorated. Nevertheless we encourage the use of cenH3, especially when the context is chromatin or histones, and in organisms in which orthology to animal CENP-A or fungal Cse4 is uncertain.

H3t has now been applied to both humans and urochordates, but the proteins are not orthologous [28]. We recommend using the descriptor TS before these: TS H3.4 (the original name of human H3t) [8] in humans and TS H3.4.1 to H3.4.3 in urochordates (no orthology with human H3.4 implied) [28]. H2Bv (or H2BV) has been used in *Plasmodium*, *Toxoplasma* and *Trypanosoma* [20-22]. In *Plasmodium* and *Toxoplasma*, the two H2Bvs are apparent orthologs, but this is unlikely to be the case for the *Trypanosoma* variant (Figure 2). Priority

for the name H2BV goes to *Trypanosoma* [21], which also has H3V and H4V [22]. We suggest these be reformatted to H2B.V, H3.V, and H4.V (Figure 4). In *Toxoplasma*, H2Bv is associated with H2A.Z [68] (which

intriguingly is also true in *Trypanosoma* H2BV), suggesting that an alternative name that does not imply identity with the *Trypanosoma* variant might be H2B.Z. H3v has been used with paralog numbers (H3v1 to H3v10) for



the many H3 variants in *Stylonychia* [26,27]. We suggest that the 'v' in these names be replaced with a period (H3.1 to H3.10), in keeping with other paralog number designations.

Additional considerations

As with 'H3', we encourage the use of 'H2A' and 'H2B' in the inclusive sense, and the use of a descriptor, such as 'RC', to specify replication-coupled forms. 'H2A.X' has traditionally been used to designate the subset of H2A variants that bear a terminal SQ(E/D)Φ phosphorylation motif, often used in contrast to unqualified "H2A", as shown in Table 3. Phylogenetic analysis indicates that SQ(E/D)Φ-bearing variants have diverged repeatedly from variants lacking this motif [47]. This makes the designation of SQ(E/D)Φ-bearing variants as 'H2A.X' either paraphyletic or polyphyletic, depending on whether the motif is ancestral or not. Despite our desire to use letter suffixes for monophyletic clades, we see no easy solution to this other than to continue with historical usage. The situation is analogous to that with H3.3, except that we allow number suffixes to be organism-specific without phylogenetic implication.

H2A.Bbd [13] (or as we would prefer, H2A.B) is a growing subfamily of histones that appears to be related to the H2AL (or H2A.L) subfamily [24]. Both families are involved in mammalian spermiogenesis [24,69,70], and are rapidly evolving, with lineage specific expansions (Figure 1). Both have shortened docking domains and wrap less DNA than other H2As [14,71]. H2A.L forms subnucleosomal particles with TH2B (TS H2B.1) [24]. When these two H2A subfamilies are better understood, there may be a logical method of combining them with a prefix for short wrapping (or another characteristic); however, they represent distinct clades, and both are widely distributed in mammals, so we conservatively recommend treating them as distinct variants at the present.

In general, many new variants are likely to be testis-specific or pollen-specific, and we urge caution and conservatism in assigning them new names. Indications are that these variants are common, polyphyletic, rapidly evolving and may have unusual properties. Orthologies and paralogies may be difficult to disentangle. We recommend naming variants initially with paralog numbers and then renaming them when their properties and relationships are better understood. We encourage the use of descriptors (for example, TS) rather than prefixes and letter suffixes when only tissue- or cell-specific expression patterns distinguish these variants from other similar variants.

Nomenclature for histone H1

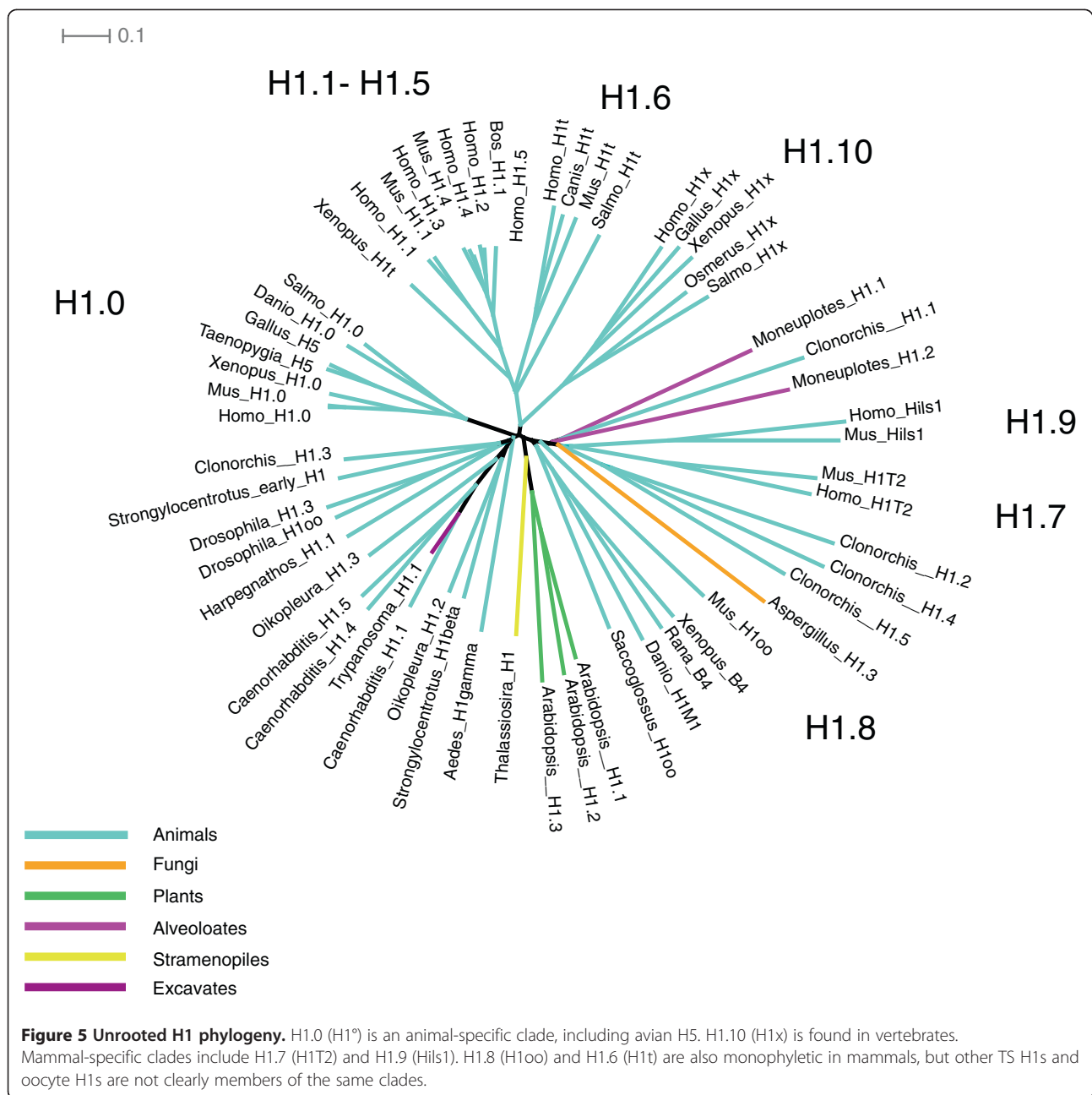
Histone H1 differs dramatically from the core histones. It has an entirely separate origin, probably from bacterial

proteins rather than from archaeal histones [72]. Rather than a histone fold, H1s typically have a short basic amino-terminal domain, a globular winged-helix domain and a lysine-rich carboxy-terminal domain often characterized by a proline-kinked alanine-lysine helix. H1s are less conserved than other histones. In some unicellular eukaryotes, such as Euglenozoa and Alveolata, H1s lack the winged-helix domain.

Most studies of H1 have taken place in animals. The discovery of H1 variants in calf thymus preceded the discovery of core histone variants by several years [73,74], and 11 variants have now been identified in humans [75,76]. As with core histones, a variety of naming styles have been applied to H1s in different organisms, including paralog numbers, letter suffixes, and combination letter and number suffixes. An early and widely-used nomenclature used lower case letters to designate paralogs in the order of elution from a Bio-Rex 70 column [77], and was subsequently adopted for variants separated on 2-D gels [78,79]. Confusion over some 12 different nomenclatures led to a previous attempt to create a system in which variant designations were applied uniformly to orthologs across mammalian species [80].

The cloning of human H1 genes introduced a nomenclature that more closely resembles core histone names in the use of a period before a paralog number [81-83], and that is now commonly used for human variants. Human H1s are often subdivided by the use of descriptors into somatic H1s and germ cell H1s. The somatic H1s include H1.1 to H1.5, H1x, and H1° or H1.0. Germ cell H1s have been designated H1t, H1T2, Hils1 (all testis-specific) and H1oo (oocyte-specific). H1s in other less well-characterized organisms are designated with paralog numbers, lower case letters, or even Greek letters. Can the phylogenetic approach and conventions proposed here for the core histones be of help for standardizing H1 nomenclature?

Constructing a phylogeny of H1s yields a 'star' pattern with long branches converging on a center that has low resolution of branching (Figure 5). The short branches in the center are unstable with respect to the choice of the substitution matrix used to construct the tree. Several animal-specific lineages appear to be as distantly related to each other as they are to lineages in other eukaryotes. While the underrepresentation of most eukaryotic groups contributes to the poorly resolved branching, the larger factor is likely to be the relatively faster rate of evolution for H1s, especially germ cell H1s, over core histones, leaving distant homologs too diverged to construct an informative tree. The sheer number of H1 variants places constraints on naming them, because with 11 variants just in humans, it is easy to foresee running out of single letters to use as suffixes. Thus the default for discovering new H1s should be to assign them



organism-specific paralog numbers, since distant orthologies that would support a letter suffix are unlikely to be verifiable.

Mammals present some exceptions to this lack of detectable orthology. The human H1 variants H1.1 to H1.5 form a clade, and the individual variants have orthologs in other mammalian species, which can be clearly identified by their gene organization as well as their sequence [6]. However, these orthologs in different species have not always used the same nomenclature (for example, human H1.4 vs. mouse H1e), nor been assigned the same paralog numbers (for example, bovine H1.1 is not

orthologous to human H1.1). The consistent use of the same paralog numbers for orthologs in different mammals has been the goal of a unified nomenclature for over 15 years, and should be adopted. The designations based on cloned human genes use the same format as the core histones, and their adoption for orthologs in other mammals offers the possibility of a unifying nomenclature for all histones.

H1° is widespread in animals [84-88], and already has an alternate name, H1.0, that conforms to our proposed nomenclature. H5 from chicken erythrocytes was known at its discovery to be an equivalent of H1 [89,90], and

has long been known to be a specific ortholog of H1.0 [84,90,91]. Despite over 35 years of literature using 'H5', we find this name to be actively misinformative, since H5 does not form a separate high-level structural class of histones, and we recommend that it be replaced with 'H1.0', mentioning H5 as a synonym. A descriptor such as 'avian erythrocyte' can be added where necessary. The same nomenclature should also be applied in the case of H1.0 orthologs identified in other non-vertebrate metazoans, including the histone H1 δ from sea urchin [92] as well as RI H1 histones from bivalve molluscs [86-88]. Molecular phylogenetic analyses have revealed that these variants share a common monophyletic origin that can be traced back before the differentiation between protostomes and deuterostomes, very early in metazoan evolution.

H1x is found throughout vertebrates, but clear orthologs of the human germ cell variants H1t, H1T2, Hils1 and H1oo are restricted to mammals. Considering the possibility that every non-mammalian genome sequenced will introduce a similar number of H1s, we recommend assigning all of the H1 variants paralog numbers and doing away with letter suffixes for H1s. Table 4 lists the suggested designations for human H1s and their mammalian orthologs. The use of descriptors (for example, TS and OO) with the germ cell variants and listing of synonyms can help to ease the transition to the new nomenclature. H1x is designated by H1.10, which is intended as a convenient mnemonic for those familiar with Roman numerals. H1oo has been previously claimed to be

specifically related to amphibian B4/H1M and cleavage stage H1 of sea urchins [95], but their orthology appears uncertain (Figure 5). Thus, we conservatively suggest that B4 become H1.4 of *Xenopus* (no implied orthology to mammalian H1.4) rather than assuming orthology to H1oo. The use of descriptors 'oocyte' or 'OO', 'maternal', 'embryonic', 'cleavage stage' or other can be used to speak collectively of functionally similar histones in diverse animals without implying orthology.

Conclusions

We describe here a unified nomenclature for histones that is readily machine-searchable and uses a single standardized form of punctuation to delimit variants. This nomenclature encourages the use of the histone names H1, H2A, H2B, H3 and H4 to represent histone families, and the specification of particular variants within those families by the use of suffixes and a few prefixes. The variant designations incorporate phylogenetic information by attempting to restrict the use of prefixes and letter suffixes to represent monophyletic clades, with the exception of H2A.X and possibly cenH3, which designations are defined by established usage for highly conserved and clearly demarcated functions. For simplicity, we encourage the use of single letter suffixes, and recommend capitalization to be consistent with the Brno nomenclature for modifications.

This nomenclature system allows flexibility and agnosticism with regard to phylogeny of variants through the long-established use of paralog number suffixes to

Table 4 Unified nomenclature and synonyms for mammalian H1 variants

Histone Gene Cluster 1						
Human			Other Mammals			New unified
Gene symbol	Albig and Doenecke [83]	Ohe and Iwai [93]	mouse gene symbol	Parsegian and Hamkalo [94]	Seyedin and Kistler, Lennox and Cohen [77,78]	
<i>HIST1H1A</i>	H1.1		<i>Hist1h1a</i>	H1a	H1a	H1.1
<i>HIST1H1B</i>	H1.5	H1a	<i>Hist1h1b</i>	H1 ^s -3	H1b	H1.5
<i>HIST1H1C</i>	H1.2	H1d	<i>Hist1h1c</i>	H1 ^s -1	H1c	H1.2
<i>HIST1H1D</i>	H1.3	H1c	<i>Hist1h1d</i>	H1 ^s -2	H1d	H1.3
<i>HIST1H1E</i>	H1.4	H1b	<i>Hist1h1e</i>	H1 ^s -4	H1e	H1.4
<i>HIST1H1T</i>	H1t		<i>Hist1h1t</i>		H1t	(TS) H1.6

Orphan Genes						
Human			Mouse		New unified	
gene symbol	alias	full name	gene symbol	alias		
<i>H1FO</i>	H1.0, H1 ^o	H1 histone family, member 0	<i>H1fo</i>	H1(0)	H1.0	
<i>H1FNT</i>	H1T2	H1 histone family, member N, testis-specific	<i>H1fnt</i>	H1t2	(TS) H1.7	
<i>H1FOO</i>	H1oo	H1 histone family, member O, oocyte-specific	<i>H1foo</i>	H1oo	(OO) H1.8	
<i>HILS1</i>		Histone H1-like protein in spermatids 1	<i>Hils1</i>	TISP64	(TS) H1.9	
<i>H1FX</i>	H1x	H1 histone family, member X	<i>H1fx</i>	H1X	H1.10	

Table 5 Application of nomenclature to *Arabidopsis*

H1	Gene	Protein	Former name
At1g06760	<i>HON1</i>	H1.1	H1
At2g30620	<i>HON2</i>	H1.2	"
At2g18050	<i>HON3</i>	H1.3	"
H2A	Gene	Protein	Former name
At5g54640	<i>HTA1</i>	H2A.1	Canonical H2A
At4g27230	<i>HTA2</i>	H2A.2	"
At1g51060	<i>HTA10</i>	H2A.10	"
At3g20670	<i>HTA13</i>	H2A.13	"
At1g54690	<i>HTA3</i>	H2A.X.3	H2A.X
At1g08880	<i>HTA5</i>	H2A.X.5	"
At5g59870	<i>HTA6</i>	H2A.W.6	SPKK-bearing H2As
At5g27670	<i>HTA7</i>	H2A.W.7	"
At5g02560	<i>HTA12</i>	H2A.W.12	"
At2g38810	<i>HTA8</i>	H2A.Z.8	H2A.Z
At1g52740	<i>HTA9</i>	H2A.Z.9	"
At3g54560	<i>HTA11</i>	H2A.Z.11	"
At4g13570	<i>HTA4</i>	H2A.Z.4	none
H2B	Gene	Protein	Former name
At1g07790	<i>HTB1</i>	H2B.1	H2B
At5g22880	<i>HTB2</i>	H2B.2	"
At2g28720	<i>HTB3</i>	H2B.3	"
At5g59910	<i>HTB4</i>	H2B.4	"
At2g37470	<i>HTB5</i>	H2B.5	"
At3g53650	<i>HTB6</i>	H2B.6	"
At3g09480	<i>HTB7</i>	H2B.7	"
At1g08170	<i>HTB8</i>	H2B.8	"
At3g45980	<i>HTB9</i>	H2B.9	"
At5g02570	<i>HTB10</i>	H2B.10	"
At3g46030	<i>HTB11</i>	H2B.11	"
H3	Gene	Protein	Former name
At5g65360	<i>HTR1</i>	H3.1	H3.1
At1g09200	<i>HTR2</i>	"	"
At3g27360	<i>HTR3</i>	"	"
At5g10400	<i>HTR9</i>	"	"
At5g10390	<i>HTR13</i>	"	"
At4g40030	<i>HTR4</i>	H3.3	H3.3
At4g40040	<i>HTR5</i>	"	"
At5g10980	<i>HTR8</i>	"	"
At1g13370	<i>HTR6</i>	H3.6	none
At1g19890	<i>HTR10</i>	H3.10	MGH3/HTR10
At1g75600	<i>HTR14</i>	H3.14	none
At1g01370	<i>HTR12</i>	cenH3	CENH3/CENP-A/HTR12

Table 5 Application of nomenclature to *Arabidopsis*
(Continued)

At5g65350	<i>HTR11</i>	H3.11	none
At5g12910	<i>HTR15</i>	H3.15	"
At1g75610	<i>HTR7</i>	H3.7	"
H4	Gene	Protein	Former name
At3g46320	<i>HF01</i>	H4	H4
At5g59690	<i>HF02</i>	"	"
At2g28740	<i>HF03</i>	"	"
At1g07820	<i>HF04</i>	"	"
At3g53730	<i>HF05</i>	"	"
At5g59970	<i>HF06</i>	"	"
At3g45930	<i>HF07</i>	"	"
At1g07660	<i>HF08</i>	"	"

In *Arabidopsis*, each H1, H2A and H2B variant has a unique sequence encoded by a unique gene. Multiple genes encode H4 and some H3 variants. H2A and H2B variants can be grouped into clades of more closely related variants. For H2A, these have distinct structures and properties that are marked by letter suffixes: H2A.W, H2A.X and H2A.Z. Individual variants within these clades are indicated by the addition of a paralog number. For H2B, any differences in properties between clades are not yet understood, and only paralog numbers are used to specify variants: H2B.1 to H2B.11. As a matter of mnemonic convenience, paralog numbers are chosen to match pre-existing gene numbers in ChromDB [64], except for those variants (H3.1, H3.3 and H4) encoded by more than one gene.

indicate individual unique variants on an organism-specific basis without implying phylogenetic relationships. Where orthologies are clear between related numbered variants, as in mammalian H1s, we encourage the adoption of paralog numbers that are consistent with known orthologs in related organisms, though paralog numbers should not be assumed to be orthologous between species without specific knowledge. The assignment of paralog numbers within an organism need not be sequential if a mnemonic or other purpose is served by choosing otherwise, for example, to bring gene and protein numbers into conformity. Consistent with current usage, we suggest reserving the designation H3.3 for the major replication-independent or replacement H3 in a particular organism. The system is adaptable to include information on multiple steps of phylogenetic divergence through treating each period (.) as a phylogenetic branchpoint, as in vertebrate H2A.Z.1 and H2A.Z.2. Splice variants should be treated like other paralogs, but can optionally be indicated by the addition of 's' before a paralog number where desirable for clarity, as in macroH2A.1.s1 and macroH2A.1.s2.

Our system has attempted to accommodate historical usage where it does not conflict with the underlying principles and, in a few cases, where it does conflict. Where existing names are changed by our guidelines, we

strongly recommend that authors include both the old and new names in the abstract of their reports to facilitate literature searches. We encourage the use of descriptors for specifying species, functional properties and tissue- or stage-specific expression. Such descriptors are intentionally not standardized to assure flexibility, though some descriptors, such as testis-specific (TS), may become commonly used. We recommend that each new histone variant by default be assigned a paralog number, with a letter suffix assigned only if helpful to call out distinctive families of variants as phylogeny and protein properties become clear. The use of new prefixes should meet an even higher standard of need and significance. An example of how to apply these guidelines to a particular organism is given in Table 5 for *Arabidopsis*.

Abbreviations

CS: Cleavage stage; GC: Generative cell-specific; OO: Oocyte-specific; RC: Replication-coupled; RI: Replication-independent; TS: Testis-specific.

Competing interests

The authors declare no competing interests.

Acknowledgments

This project was initiated during the EMBO Workshop on Histone variants and genome regulation held in Strasbourg, France, 12–14 October 2011. We are grateful for support from EMBO, the Howard Hughes Medical Institute, and the Intramural Research Program of the NIH, National Library of Medicine.

Author details

¹Howard Hughes Medical Institute, Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. ²Department of BCMP, Harvard Medical School, Boston, MA 02115, USA. ³CNRS, UMR 218, Institut Curie, Centre de Recherche, Paris, F-75248 cx 05, France. ⁴Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8W 3P6, Canada. ⁵Temasek Lifesciences Laboratory, 1 Research Link, National University of Singapore, Singapore, 117604, Singapore. ⁶Plant Molecular Biology and Biotechnology Group, Melbourne School of Land and Environment, The University of Melbourne, Parkville, VIC 3010, Australia. ⁷Laboratory of Molecular Pharmacology, CCR, NCI, NIH, Bethesda, MD 20892, USA. ⁸Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3200, USA. ⁹Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA. ¹⁰Department of Plant Biology, UC Davis / HHMI, Davis, CA 95616, USA. ¹¹Laboratory of Molecular Parasitology, The Rockefeller University, New York, NY 10065, USA. ¹²Department of Entomology, Pennsylvania State University, University Park, PA 16802, USA. ¹³Laboratoire de Biologie Moléculaire et Cellulaire de la Différenciation, Institut Albert Bonniot, INSERM/UJF U821, Grenoble, France. ¹⁴Department of Biochemistry, University of Goettingen, Goettingen, D-37073, Germany. ¹⁵Department of Cellular and Molecular Biology, University of A Coruna, A Coruna, E15071, Spain. ¹⁶Department of Biology, University of Rochester, Rochester, NY 14627, USA. ¹⁷Center for Integrated Protein Science Munich at the Adolf-Butenandt Institute, Department for Molecular Biology, Ludwig-Maximilians-University Munich, Munich, 80336, Germany. ¹⁸Department of Molecular Biology and Biochemistry, University of CA, Irvine, CA 92697, USA. ¹⁹Howard Hughes Medical Institute, Department of Molecular Cellular and Developmental Biology, University of California, Los Angeles, CA 9009, USA. ²⁰Max Planck Institute for Immunobiology and Epigenetics, Freiburg, 79108, Germany. ²¹INSERM, U823; Université Joseph Fourier - Grenoble 1, La Tronche, F-38706, France. ²²Department of Physiological Chemistry, Butenandt Institute and Biomedical Center, Ludwig Maximilians University of Munich, Munich, 81377, Germany. ²³National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894, USA. ²⁴CGphiMC - CNRS UMR5534, Claude Bernard University Lyon1 - University of Lyon, Villeurbanne, 69622 Cedex, France. ²⁵Program in Molecular Biology and Biotechnology, University of North Carolina, Chapel

Hill, NC 27599, USA. ²⁶Department of Animal Biology, University of Pennsylvania, Philadelphia, PA 19104-6046, USA. ²⁷HELIOS Medical Centre Wuppertal, Paediatrics Centre, Witten/Herdecke University, Wuppertal, D-42283, Germany. ²⁸Institut de Genetique et Biologie Moléculaire et Cellulaire, Illkirch, 67404, France. ²⁹Department of Microbiology, Immunology, and Cancer Biology, University of Virginia, Charlottesville, VA 22908, USA. ³⁰Sars International Center for Marine Molecular Biology and Department of Biology, University of Bergen, Bergen, N-5008, Norway. ³¹Stem Cells and Developmental Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM U964, Université de Strasbourg, Illkirch, CU de Strasbourg F-67404, France. ³²The John Curtin School of Medical Research, Genome Biology Department, The Australian National University, Canberra, ACT 2601, Australia. ³³College of Medical and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK. ³⁴Cell Biology and Biophysics, School of Biological Sciences, University of Missouri-Kansas City, Kansas City, MO 64110, USA. ³⁵Chromatin Lab, National Institute of Biological Sciences, Beijing, Beijing 102206, China.

Authors' contributions

PBT and SH designed the study. FB, SHo, HW and RY contributed the *Arabidopsis* tables, and RS contributed the H1 table. All other authors provided input during early discussions and drafts, and PBT wrote the paper. All authors read and approved the final manuscript.

Received: 18 February 2012 Accepted: 12 April 2012

Published: 31 May 2012

References

1. Bradbury EM: **Histone nomenclature.** *Methods Cell Biol* 1977, **16**:179–181.
2. Alfageme CR, Zweidler A, Mahowald A, Cohen LH: **Histones of *Drosophila* embryos. Electrophoretic isolation and structural studies.** *J Biol Chem* 1974, **249**:3729–3736.
3. Strickland WN, Strickland M, Brandt WF, Morgan M, Von Holt C: **Partial amino acid sequence of two new arginine-serine rich histones from male gonads of the sea urchin (*Parechinus angulosus*).** *FEBS Lett* 1974, **40**:161–166.
4. Marzluff WF Jr, Sanders LA, Miller DM, McCarty KS: **Two chemically and metabolically distinct forms of calf thymus histone F3.** *J Biol Chem* 1972, **247**:2026–2033.
5. Patthy L, Smith EL: **Histone III. VI. Two forms of calf thymus histone III.** *J Biol Chem* 1975, **250**:1919–1920.
6. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ: **The human and mouse replication-dependent histone genes.** *Genomics* 2002, **80**:487–498.
7. Franklin SG, Zweidler A: **Non-allelic variants of histones 2a, 2b and 3 in mammals.** *Nature* 1977, **266**:273–275.
8. Albig W, Ebertauer J, Klobbeck G, Kunz J, Doenecke D: **A solitary human H3 histone gene on chromosome 1.** *Hum Genet* 1996, **97**:486–491.
9. Schenk R, Jenke A, Zilbauer M, Wirth S, Postberg J: **H3.5 is a novel hominid-specific histone H3 variant that is specifically expressed in the seminiferous tubules of human testes.** *Chromosoma* 2011, **120**:275–285.
10. West MH, Bonner WM: **Histone 2A, a heteromorphous family of eight protein species.** *Biochemistry* 1980, **19**:3238–3245.
11. Witt O, Albig W, Doenecke D: **Testis-specific expression of a novel human H3 histone gene.** *Exp Cell Res* 1996, **229**:301–306.
12. Ghosh S, Klobutcher LA: **A development-specific histone H3 localizes to the developing macronucleus of *Euplotes*.** *Genesis* 2000, **26**:179–188.
13. Chadwick BP, Willard HF: **A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome.** *J Cell Biol* 2001, **152**:375–384.
14. Bao Y, Konesky K, Park YJ, Rosu S, Dyer PN, Rangasamy D, Tremethick DJ, Laybourn PJ, Luger K: **Nucleosomes containing the histone variant H2A. Bbd organize only 118 base pairs of DNA.** *EMBO J* 2004, **23**:3314–3324.
15. Boulard M, Gautier T, Mbele GO, Gerson V, Hamiche A, Angelov D, Bouvet P, Dimitrov S: **The NH2 tail of the novel histone variant H2BFWT exhibits properties distinct from conventional H2B with respect to the assembly of mitotic chromosomes.** *Mol Cell Biol* 2006, **26**:1518–1526.
16. Van Doninck K, Mandigo ML, Hur JH, Wang P, Guglielmini J, Milinkovitch MC, Lane WS, Meselson M: **Phylogenomics of unusual histone H2A Variants in *Bdelloid* rotifers.** *PLoS Genet* 2009, **5**:e1000401.
17. Wiedemann SM, Mildner SN, Bonisch C, Israel L, Maiser A, Matheisl S, Straub T, Merkl R, Leonhardt H, Kremmer E, Schermelleh L, Hake SB: **Identification**

- and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y. *J Cell Biol* 2010, **190**:777–791.
18. van Daal A, Elgin SC: A histone variant, H2AvD, is essential in *Drosophila melanogaster*. *Mol Biol Cell* 1992, **3**:593–602.
 19. Leach TJ, Mazzeo M, Chotkowski HL, Madigan JP, Wotring MG, Glaser RL: Histone H2A.Z is widely but nonrandomly distributed in chromosomes of *Drosophila melanogaster*. *J Biol Chem* 2000, **275**:23267–23272.
 20. Miao J, Fan Q, Cui L, Li J, Li J, Cui L: The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene* 2006, **369**:53–65.
 21. Lowell JE, Kaiser F, Janzen CJ, Cross GA: Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in *Trypanosoma brucei*. *J Cell Sci* 2005, **118**:5721–5730.
 22. Siegel TN, Hekstra DR, Kemp LE, Figueiredo LM, Lowell JE, Fenyo D, Wang X, Dewell S, Cross GA: Four histone variants mark the boundaries of polycistronic transcription units in *Trypanosoma brucei*. *Genes Dev* 2009, **23**:1063–1076.
 23. Dawson SC, Sagolla MS, Cande WZ: The cenH3 histone variant defines centromeres in *Giardia intestinalis*. *Chromosoma* 2007, **116**:175–184.
 24. Govin J, Escoffier E, Rousseaux S, Kuhn L, Ferro M, Thevenon J, Catena R, Davidson I, Garin J, Khochbin S, Caron C: Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *J Cell Biol* 2007, **176**:283–294.
 25. Eirin-Lopez JM, Gonzalez-Romero R, Dryhurst D, Ishibashi T, Ausio J: The evolutionary differentiation of two histone H2A.Z variants in chordates (H2A.Z-1 and H2A.Z-2) is mediated by a stepwise mutation process that affects three amino acid residues. *BMC Evol Biol* 2009, **9**:31.
 26. Bernhard D: Several highly divergent histone H3 genes are present in the hypotrichous ciliate *Stylonychia lemnae*. *FEMS Microbiol Lett* 1999, **175**:45–50.
 27. Postberg J, Forcob S, Chang WJ, Lipps HJ: The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol Biol* 2010, **10**:259.
 28. Moosmann A, Campsteijn C, Jansen PW, Nasrallah C, Raasholm M, Stunnenberg HG, Thompson EM: Histone variant innovation in a rapidly evolving chordate lineage. *BMC Evol Biol* 2011, **11**:208.
 29. Pehrson JR, Costanzi C, Dharia C: Developmental and tissue expression patterns of histone macroH2A1 subtypes. *J Cell Biochem* 1997, **65**:107–113.
 30. Brock WA, Trostle PK, Meistrich ML: Meiotic synthesis of testis histones in the rat. *Proc Natl Acad Sci U S A* 1980, **77**:371–375.
 31. Zalensky AO, Siino JS, Gineitis AA, Zalenskaya IA, Tomilin NV, Yau P, Bradbury EM: Human testis/sperm-specific histone H2B (hTSH2B). Molecular cloning and characterization. *J Biol Chem* 2002, **277**:43474–43480.
 32. Ueda K, Tanaka I: The appearance of male gamete-specific histones gH2B and gH3 during pollen development in *Lilium longiflorum*. *Dev Biol* 1995, **169**:210–217.
 33. Xu H, Swoboda I, Bhalla PL, Singh MB: Male gametic cell-specific expression of H2A and H3 histone genes. *Plant Mol Biol* 1999, **39**:607–614.
 34. Okada T, Singh MB, Bhalla PL: Histone H3 variants in male gametic cells of lily and H3 methylation in mature pollen. *Plant Mol Biol* 2006, **62**:503–512.
 35. Pehrson JR, Fried VA: MacroH2A, a core histone containing a large nonhistone region. *Science* 1992, **257**:1398–1400.
 36. Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S: Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 2002, **14**:1053–1066.
 37. Aul RB, Oko RJ: The major subsacrosomal occupant of bull spermatozoa is a novel histone H2B variant associated with the forming acrosome during spermiogenesis. *Dev Biol* 2002, **242**:376–387.
 38. Marzluff WF, Sakallah S, Kelkar H: The sea urchin histone gene complement. *Dev Biol* 2006, **300**:308–320.
 39. Turner BM: Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat Struct Mol Biol* 2005, **12**:110–112.
 40. Dryhurst D, Ishibashi T, Rose KL, Eirin-Lopez JM, McDonald D, Silva-Moreno B, Veldhoen N, Helbing CC, Hendzel MJ, Shabanowitz J, Hunt DF, Ausio J: Characterization of the histone H2A.Z-1 and H2A.Z-2 isoforms in vertebrates. *BMC Biol* 2009, **7**:86.
 41. Nei M, Rooney AP: Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 2005, **39**:121–152.
 42. Ingouff M, Hamamura Y, Gourgues M, Higashiyama T, Berger F: Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. *Curr Biol* 2007, **17**:1032–1037.
 43. Ferguson L, Ellis PJ, Affara NA: Two novel mouse genes mapped to chromosome Yp are expressed specifically in spermatids. *Mamm Genome* 2009, **20**:193–206.
 44. Talbert PB, Henikoff S: Histone variants—ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* 2010, **11**:264–275.
 45. Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R: Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 2007, **8**:460.
 46. Baker RE, Rogers K: Phylogenetic analysis of fungal centromere H3 proteins. *Genetics* 2006, **174**:1481–1492.
 47. Malik HS, Henikoff S: Phylogenomics of the nucleosome. *Nat Struct Biol* 2003, **10**:882–891.
 48. Talbert PB, Bayes JJ, Henikoff S: The Evolution of Centromeres and Kinetochores: A Two-Part Fugue. In *The Kinetochores: From Molecular Discoveries to Cancer Therapy*. 1st edition. Edited by De Wulf P, Earnshaw WC. Berlin: Springer; 2008:193–230.
 49. Tran MH, Aul RB, Xu W, van der Hoorn F, Oko R: Involvement of Classical Bipartite/Karyopherin Nuclear Import Pathway Components in Acrosomal Trafficking and Assembly During Bovine and Murid Spermiogenesis. *Biol Reprod* 2012, **86**:84.
 50. Lindsey GG, Orgeig S, Thompson P, Davies N, Maeder DL: Extended C-terminal tail of wheat histone H2A interacts with DNA of the "linker" region. *J Mol Biol* 1991, **218**:805–813.
 51. Lee J, Park HS, Kim HH, Yun YJ, Lee DR, Lee S: Functional polymorphism in H2BFWT -5'UTR is associated with susceptibility to male infertility. *J Cell Mol Med* 2009, **13**:1942–1951.
 52. Song X, Bowen J, Miao W, Liu Y, Gorovsky M: The non-histone, N-terminal tail of an essential, chimeric H2A variant regulates mitotic H3 S10 de-phosphorylation. *Genes Dev* 2012, **26**:615–629.
 53. Costanzi C, Pehrson JR: MACROH2A2, a new member of the MARCOH2A core histone family. *J Biol Chem* 2001, **276**:21776–21784.
 54. Waterborg JH: Dynamics of histone acetylation *in vivo*. A function for acetylation turnover? *Biochem Cell Biol* 2002, **80**:363–378.
 55. Waterborg JH: Evolution of histone H3: emergence of variants and conservation of post-translational modification sites. *Biochem Cell Biol* 2012, **90**:79–95.
 56. Thatcher TH, MacGaffey J, Bowen J, Horowitz S, Shapiro DL, Gorovsky MA: Independent evolutionary origin of histone H3.3-like variants of animals and *Tetrahymena*. *Nucleic Acids Res* 1994, **22**:180–186.
 57. Waterborg JH, Robertson AJ: Common features of analogous replacement histone H3 genes in animals and plants. *J Mol Evol* 1996, **43**:194–206.
 58. Chaubet N, Clement B, Gigot C: Genes encoding a histone H3.3-like variant in *Arabidopsis* contain intervening sequences. *J Mol Biol* 1992, **225**:569–574.
 59. Borun TW, Ajiro K, Zweidler A, Dolby TW, Stephens RE: Studies of human histone messenger RNA. II. The resolution of fractions containing individual human histone messenger RNA species. *J Biol Chem* 1977, **252**:173–180.
 60. Cui B, Liu Y, Gorovsky MA: Deposition and function of histone H3 variants in *Tetrahymena thermophila*. *Mol Cell Biol* 2006, **26**:7719–7730.
 61. Sullivan WJ Jr: Histone H3 and H3.3 variants in the protozoan pathogens *Plasmodium falciparum* and *Toxoplasma gondii*. *DNA Seq* 2003, **14**(3):227–231.
 62. Akhmanova AS, Bindels PC, Xu J, Miedema K, Kremer H, Hennig W: Structure and expression of histone H3.3 genes in *Drosophila melanogaster* and *Drosophila hydei*. *Genome* 1995, **38**:586–600.
 63. Rando OJ: Genome-wide measurement of histone H3 replacement dynamics in yeast. *Methods Mol Biol* 2011, **759**:41–60.
 64. Chrom DB: The Chromatin Database <http://www.chromdb.org>.
 65. Okada T, Endo M, Singh MB, Bhalla PL: Analysis of the histone H3 gene family in *Arabidopsis* and identification of the male-gamete-specific variant AtMGH3. *Plant J* 2005, **44**:557–568.
 66. Earnshaw WC, Rothfield N: Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 1985, **91**:313–321.
 67. Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M: 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* 1995, **9**:573–586.

68. Dalmaso MC, Onyango DO, Naguleswaran A, Sullivan WJ Jr, Angel SO: **Toxoplasma H2A variants reveal novel insights into nucleosome composition and functions for this histone family.** *J Mol Biol* 2009, **392**:33–47.
69. Ishibashi T, Li A, Eirin-Lopez JM, Zhao M, Missiaen K, Abbott DW, Meistrich M, Hendzel MJ, Ausio J: **H2A.Bbd: an X-chromosome-encoded histone involved in mammalian spermiogenesis.** *Nucleic Acids Res* 2010, **38**:1780–1789.
70. Soboleva TA, Nekrasov M, Pahwa A, Williams R, Huttley GA, Tremethick DJ: **A unique H2A histone variant occupies the transcriptional start site of active genes.** *Nat Struct Mol Biol* 2011, **19**:25–30.
71. Syed SH, Boulard M, Shukla MS, Gautier T, Travers A, Bednar J, Faivre-Moskalenko C, Dimitrov S, Angelov D: **The incorporation of the novel histone variant H2AL2 confers unusual structural and functional properties of the nucleosome.** *Nucleic Acids Res* 2009, **37**:4684–4695.
72. Kasinsky HE, Lewis JD, Dacks JB, Ausio J: **Origin of H1 linker histones.** *FASEB J* 2001, **15**:34–42.
73. Kinkade JM Jr, Cole RD: **The resolution of four lysine-rich histones derived from calf thymus.** *J Biol Chem* 1966, **241**:5790–5797.
74. Kinkade JM Jr, Cole RD: **A structural comparison of different lysine-rich histones of calf thymus.** *J Biol Chem* 1966, **241**:5798–5805.
75. Happel N, Doenecke D: **Histone H1 and its isoforms: contribution to chromatin structure and function.** *Gene* 2009, **431**:1–12.
76. Izzo A, Kamieniarz K, Schneider R: **The histone H1 family: specific members, specific functions?** *Biol Chem* 2008, **389**:333–343.
77. Seyedin SM, Kistler WS: **H1 histone subfractions of mammalian testes. 1. Organ specificity in the rat.** *Biochemistry* 1979, **18**:1371–1375.
78. Lennox RW, Cohen LH: **The histone H1 complements of dividing and nondividing cells of the mouse.** *J Biol Chem* 1983, **258**:262–268.
79. Lennox RW, Cohen LH: **The alterations in H1 histone complement during mouse spermatogenesis and their significance for H1 subtype function.** *Dev Biol* 1984, **103**:80–84.
80. Parseghian MH, Henschen AH, Krieglstein KG, Hamkalo BA: **A proposal for a coherent mammalian histone H1 nomenclature correlated with amino acid sequences.** *Protein Sci* 1994, **3**:575–587.
81. Albig W, Kardalidou E, Drabent B, Zimmer A, Doenecke D: **Isolation and characterization of two human H1 histone genes within clusters of core histone genes.** *Genomics* 1991, **10**:940–948.
82. Albig W, Drabent B, Kunz J, Kalf-Suske M, Grzeschik KH, Doenecke D: **All known human H1 histone genes except the H1(0) gene are clustered on chromosome 6.** *Genomics* 1993, **16**:649–654.
83. Albig W, Kioschis P, Poustka A, Meergans K, Doenecke D: **Human histone gene organization: nonregular arrangement within a large cluster.** *Genomics* 1997, **40**:314–322.
84. Peretti M, Khochbin S: **The evolution of the differentiation-specific histone H1 gene basal promoter.** *J Mol Evol* 1997, **44**:128–134.
85. Brocard MP, Triebe S, Peretti M, Doenecke D, Khochbin S: **Characterization of the two H1(zero)-encoding genes from *Xenopus laevis*.** *Gene* 1997, **189**:127–134.
86. Eirín-López JM, González-Tizón AM, Martínez A, Méndez J: **Molecular and evolutionary analysis of mussel histone genes (*Mytilus* spp.): possible evidence of an "orphan origin" for H1 histone genes.** *J Mol Evol* 2002, **55**:272–283.
87. González-Romero R, Ausió J, Méndez J, Eirín-López JM: **Early evolution of histone genes: prevalence of an 'orphan' H1 lineage in protostomes and birth-and-death process in the H2A family.** *J Mol Evol* 2008, **66**:505–518.
88. González-Romero R, Ausió J, Méndez J, Eirín-López JM: **Histone genes of the razor clam *Solen marginatus* unveil new aspects of linker histone evolution in protostomes.** *Genome* 2009, **52**:597–607.
89. Garel A, Mazen A, Champagne M, Sautiere P, Kmiecik D, Loy O, Biserte G: **Chicken erythrocyte histone H5; I. Amino terminal sequence (70 residues).** *FEBS Lett* 1975, **50**:195–199.
90. Khochbin S: **Histone H1 diversity: bridging regulatory signals to linker histone function.** *Gene* 2001, **271**:1–12.
91. Smith BJ, Walker JM, Johns EW: **Structural homology between a mammalian H1(0) subfraction and avian erythrocyte-specific histone H5.** *FEBS Lett* 1980, **112**:42–44.
92. Lieber T, Angerer LM, Angerer RC, Childs G: **A histone H1 protein in sea urchins is encoded by a poly(A) + mRNA.** *Proc Natl Acad Sci U S A* 1988, **85**:4123–4127.
93. Ohe Y, Hayashi H, Iwai K: **Human spleen histone H1. Isolation and amino acid sequences of three minor variants, H1a, H1c, and H1d.** *J Biochem* 1989, **106**:844–857.
94. Parseghian MH, Clark RF, Hauser LJ, Dvorkin N, Harris DA, Hamkalo BA: **Fractionation of human H1 subtypes and characterization of a subtype-specific antibody exhibiting non-uniform nuclear staining.** *Chromosome Res* 1993, **1**:127–139.
95. Tanaka M, Hennebold JD, Macfarlane J, Adashi EY: **A mammalian oocyte-specific linker histone gene H1oo: homology with the genes for the oocyte-specific cleavage stage histone (cs-H1) of sea urchin and the B4/H1M histone of the frog.** *Development* 2001, **128**:655–664.

doi:10.1186/1756-8935-5-7

Cite this article as: Talbert *et al.*: A unified phylogeny-based nomenclature for histone variants. *Epigenetics & Chromatin* 2012 **5**:7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

