Cell PRESS

Phylogeny as the basis for naming histones

Paul B. Talbert and Steven Henikoff

Howard Hughes Medical Institute and Basic Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA

Thirty years ago, most proteins were still discovered by protein sequencing, whereas in the genomic era, most proteins are now discovered by conceptually translating DNA sequences, and many are found to be members of protein families with orthologs and paralogs in multiple organisms. The naming of members of large protein families can rapidly become haphazard or contradictory; therefore, nomenclature revisions and rules are often sought by researchers to prevent confusion. Such has been the case for histones, for which naming rules had been established in 1977, based largely on their biophysical properties [1]. As part of a community effort [2], we recently updated and systematized the nomenclature for histones following the logic of phylogeny, which has been used in other nomenclature revisions because it encapsulates protein history and helps to predict structure. In contrast to previous such efforts, which had resulted in wholesale renaming, for example of kinesins [3], we retained the 1977 protein family names: H1, H2A, H2B, H3, and H4. We also retained the traditional period (.) punctuation for indicating variants (e.g., H3.1, H3.2, and H3.3), extending its use to denote phylogenetic branch points at all levels, including alternatively processed transcripts. To help illustrate the reasoning for our unified nomenclature system, we included comprehensive phylogenies for all five histone families [2].

The use of phylogeny as the basis for protein family nomenclature is so compelling and widely accepted that it is a surprise to learn that our guidelines have met with resistance from a large fraction of the centromere community, even though less than 2% of our text concerned centromere-specific histones. In a recent publication, Earnshaw *et al.* [4] likened our revisions to 'Esperanto', suggesting that we had disregarded conventions of naming priority by attempting to supplant the name 'CENP-A' (centromere protein A) with 'CenH3', and that all H3 histones at centromeres should be referred to as CENP-A (Box 1). On the contrary, the 27 histone names that we had proposed changing did not include CENP-A or any of the other established centromeric histone names.

The arguments of Earnshaw *et al.* also ignored the evidence of phylogeny. Specifically, centromere-specific H3 variants do not constitute a subfamily that excludes noncentromeric H3 variants, and studies by several groups have failed to produce evidence that known centromere-specific histones are monophyletic (e.g., [5]). High bootstrap values support the monophyly of CENP-A orthologs in Amniotes and, to a lesser extent, in other vertebrates and some invertebrates, but they do not support, for example, the hypothesis that the centromere-specific H3s of *Saccharomyces* (Cse4), *Caenorhabditis* (HCP-3), *Drosophila* (Cid), or *Arabidopsis* (HTR12) are monophyletic with human CENP-A to the exclusion of other H3s [5]. It was for this reason that the term 'cenH3' [6] was introduced as a convenient way to specify the functional class of all centromere-specific H3 variants, whether or not they are monophyletic with human CENP-A. The name 'CENP-A' could then be used in the usual way (not supplanted) to designate the H3 subfamily of known orthologs to human CENP-A. By the same logic, we recognized the popularity of previously established names for other orthologous groups, including Cse4 (for fungi), Cid (for insects), and CENH3 (for plants) [2]. We also encouraged the use of cenH3 when the context is chromatin or histones, and where orthology to animal CENP-A or fungal Cse4 is uncertain.

The use of cenH3 to designate a functional class rather than a subfamily resembles the use of 'canonical H3' to designate H3 paralogs whose expression is coupled to replication, even though paralogs with this expression pattern most likely evolved independently in plants and in animals. If improved taxon sampling or advances in phylogenetic methods in the future resolve the phylogeny of all cenH3s in favor of a monophyletic and exclusively centromere-specific CENP-A subfamily of the H3 family, then the name 'cenH3' would serve as an alternate synonymous designation that specifies clearly that CENP-A belongs to the well-established H3 protein family. This is the same relation that the name 'kinesin-7' [3] has to 'CENP-E'.

Earnshaw *et al.* also raised concern that 'confusion will inevitably arise over whether the term CenH3 refers to canonical histone H3 interspersed with CENP-A at centromeres or to the CENP-A itself' in regional centromeres. The notion that the same H3 paralog would have a different name when it occurs in a particular chromosomal location is unprecedented in our experience. Moreover, by Earnshaw *et al.*'s reasoning, the CENP nomenclature (Box 1), in which more than 20 unrelated proteins are distinguished by a single letter suffix, would also be confusing. In the case of cenH3 and canonical H3, no confusion arises in practice, because cenH3 is already cross-referenced with CENP-A by PubMed so that both terms fetch the same publications.

This controversy over naming histones reflects a difference of opinion as to the importance of accepted evolutionary principles based on phylogenetic evidence by workers in the field, and parallels a recent debate about the findings of the ENCODE project [7]. In that case, evolutionary biologists disputed the claim by the ENCODE community that 80% of the human genome is functional, and marshaled compelling

Corresponding author: Henikoff, S. (steveh@fhcrc.org).
Keywords: CENP-A; cenH3; nomenclature; evolution.

Box 1. CENP-A identification and classification as a histone

The human CENP-A protein was first unequivocally identified on a western blot as a centromere-specific protein reacting to CREST antibodies by Guldner *et al.* [9], who simply called it a '19.5 kD non-histone chromosomal protein'. In January 1985, Palmer and Margolis [10] showed that a CREST-reactive centromeric protein in rats and chickens was found in mononucleosomes, and suggested that it was histone-like and substituted for a normal mononucleosome component. A month later, Earnshaw and Rothfield [11] published their identification of three human CREST-reactive proteins that they named CENP-A, CENP-B, and CENP-C, which they hypothesized formed a protein family with shared epitopes. Thus, CENP-A was correctly identified as a probable histone shortly before it was incorrectly identified and named as a member of a non-existent 'CENP' protein family.

Although bovine and human CENP-A are 78% identical and easily identified as orthologs, when the first non-vertebrate centromere-specific H3, Cse4, was identified in *Saccharomyces*, Stoler *et al.* [12] discussed its relation to H3 and CENP-A, but refrained from asserting its orthology to CENP-A or adopting the name 'CENP-A', possibly because Cse4 has greater identity to H3 than to human CENP-A (61% versus 58%).

evidence that this conclusion ignores accepted Darwinian principles [8]. Likewise, characterizing a phylogeny-based nomenclature as 'Esperanto' and ignoring the distinction between paralogy and orthology overlooks the many benefits of applying evolutionary principles to guide biological research.

References

1 Bradbury, E.M. (1977) Histone nomenclature. *Methods Cell Biol.* 16, 179–181
2 Talbert, P.B. *et al.* (2012) A unified phylogeny-based nomenclature for histone variants. *Epigenet. Chromatin* 5, 7
3 Lawrence, C.J. *et al.* (2004) A standardized kinesin nomenclature. *J. Cell Biol.* 167, 19–22
4 Earnshaw, W.C. *et al.* (2013) Esperanto for histones: CENP-A, not CenH3, is the centromeric histone H3 variant. *Chromosome Res.* 21, 101–106
5 Postberg, J. *et al.* (2010) The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evol. Biol.* 10, 259
6 Talbert, P.B. *et al.* (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14, 1053–1066
7 Dunham, I. *et al.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
8 Doolittle, W.F. (2013) Is junk DNA bunk? A critique of ENCODE. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5294–5300
9 Guldner, H.H. *et al.* (1984) Human anti-centromere sera recognise a 19.5 kD non-histone chromosomal protein from HeLa cells. *Clin. Exp. Immunol.* 58, 13–20
10 Palmer, D.K. and Margolis, R.L. (1985) Kinetochore components recognized by human autoantibodies are present on mononucleosomes. *Mol. Cell. Biol.* 5, 173–186
11 Earnshaw, W.C. and Rothfield, N. (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* 91, 313–321
12 Stoler, S. *et al.* (1995) 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

0168-9525/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.tig.2013.06.009 Trends in Genetics, September 2013, Vol. 29, No. 9

CENP-A and the CENP nomenclature: response to Talbert and Henikoff

William C. Earnshaw[1] and Don W. Cleveland[2]

[1] Wellcome Trust Centre for Cell Biology, University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JR, UK
[2] Ludwig Institute for Cancer Research, University of California at San Diego, La Jolla, CA 92093-0670, USA

Earlier this year, we, along with 55 additional investigators worldwide who with us are collectively responsible for a large majority of the literature on the components and function of centromeres and their attached kinetochores, published our recommendation [1] that the original name 'centromere protein A' ('CENP-A') continue in general use to refer to the centromeric specific variant of histone H3. This collective view was a response to a proposal by Talbert *et al.* [2] to replace the CENP-A name with a new one based solely on phylogenetic considerations.

Rather than deferring to historical precedent, and the broad consensus view of those who initially discovered and named CENP-A and those responsible for most of the subsequent relevant discoveries in species from yeast to humans, Talbert and Henikoff [3] now reiterate their earlier argument for adoption of an alternative name for this centromeric variant of histone H3. They also significantly extend their initial proposal by arguing that the initial description of CENP-A misidentified the protein as a member of what they apparently regard as a 'non-existent' CENP family (see Box 1 of their article [3]).

The logic underpinning the argument from Talbert and Henikoff is that the dominant criterion to use for determining relations between proteins should be phylogenetic. According to this view, using the term 'CENP' to specify centromere proteins is thus inappropriate, because none of the CENPs described to date (CENP-A through CENP-X) is phylogenetically related to the others. By this argument, of course, the cluster of differentiation (CD) nomenclature widely adopted to describe cell surface determinants in the immune system is also invalid, as would be the use of '$APC_{1\text{-}n}$' to describe the components of the anaphase promoting complex/cyclosome (APC/C). The list of similar examples in biology is long.

Our response is that the CENP nomenclature has been widely used since it was initially introduced in 1985 [4]

Corresponding authors: Earnshaw, W.C. (bill.earnshaw@ed.ac.uk); Cleveland, D.W. (dcleveland@ucsd.edu).
Keywords: CENP-A; kinetochore; centromere.