Research Report

Phylogeny-Based Systematization of Arabidopsis Proteins with Histone H1 Globular Domain[1[OPEN]]

Maciej Kotliński, Lukasz Knizewski, Anna Muszewska, Kinga Rutowicz, Maciej Lirski, Anja Schmidt, Célia Baroux*, Krzysztof Ginalski, and Andrzej Jerzmanowski*

Laboratory of Systems Biology, Faculty of Biology, University of Warsaw, 02-106 Warsaw, Poland (M.K., A.J.); Laboratory of Bioinformatics and Systems Biology, Centre of New Technologies, University of Warsaw, 02-089 Warsaw, Poland (L.K., K.G.); Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland (A.M., K.R., M.L., A.J.); Institute of Plant Biology and Zürich-Basel Plant Science Center, University of Zürich, 8008 Zurich, Switzerland (K.R., C.B.); and Centre for Organismal Studies, Heidelberg University, 69120 Heidelberg, Germany (A.S.)

ORCID IDs: 0000-0002-0821-077X (M.K.); 0000-0002-4578-844X (A.M.); 0000-0003-4035-675X (K.R.); 0000-0001-7831-7495 (M.L.); 0000-0001-6307-2229 (C.B.); 0000-0002-4684-4503 (A.J.).

H1 (or linker) histones are basic nuclear proteins that possess an evolutionarily conserved nucleosome-binding globular domain, GH1. They perform critical functions in determining the accessibility of chromatin DNA to trans-acting factors. In most metazoan species studied so far, linker histones are highly heterogenous, with numerous nonallelic variants cooccurring in the same cells. The phylogenetic relationships among these variants as well as their structural and functional properties have been relatively well established. This contrasts markedly with the rather limited knowledge concerning the phylogeny and structural and functional roles of an unusually diverse group of GH1-containing proteins in plants. The dearth of information and the lack of a coherent phylogeny-based nomenclature of these proteins can lead to misunderstandings regarding their identity and possible relationships, thereby hampering plant chromatin research. Based on published data and our in silico and high-throughput analyses, we propose a systematization and coherent nomenclature of GH1-containing proteins of Arabidopsis (*Arabidopsis thaliana* [L.] Heynh) that will be useful for both the identification and structural and functional characterization of homologous proteins from other plant species.

H1s, also known as linker histones, are universal and ubiquitous components of chromatin fibers, in which they occur at an average frequency of one molecule per nucleosome (Woodcock et al., 2006). They are small basic proteins with a highly conserved central globular domain (GH1) and two less conserved and mostly unstructured tail fragments: a short (~20 amino acids) N-terminal domain and a considerably longer (~100 amino acids) and highly positively charged C-terminal domain (CTD). GH1 consists of ~80 amino acids and belongs to the winged helix family of DNA-binding proteins. It contains a characteristic mixed α/β-fold consisting of three α-helices (I–III) and two β-strands (S2 and S3). The compact bundle composed of the three helices forms the core of this domain. The wing structure (from which the name of this family of DNA-binding proteins is derived) lies within the region located C terminally to helix III and is an extended loop joining β-strands S2 and S3. GH1 associates with the nucleosome outside the core particle and contacts DNA via at least two different binding sites (Zhou et al., 1998, 2013; Brown et al., 2006; Syed et al., 2010).

In addition to GH1, the overall functional properties of H1 are strongly influenced by the CTD, which binds to internucleosomal linker DNA. The CTD has an intrinsically disordered structure capable of adopting different conformations depending on the geometry of the target surfaces, which may be linker DNA or interacting proteins (Hansen et al., 2006). The prime determinant of this property is the amino acid composition rather than the CTD sequence, with charge

[1] This work was supported by the European Cooperation in Science and Technology and the Ministry of Science and Higher Education (grant no. MNiSW 212/N–COST/2008/0 to A.J. and M.K.), the Foundation for Polish Science (TEAM to K.G.) and the National Science Centre (grant nos. 2011/02/A/NZ2/00014 and 2014/15/B/NZ1/03357 to K.G.), the National Science Centre (grant no. 2012/07/D/NZ2/04286 to A.M.), the Ministry of Science and Higher Education (scholarship for outstanding young researchers to A.M.), and the Swiss National Science Foundation (grant no. 31003A_149974/1 to C.B.).

* Address correspondence to cbaroux@botinst.uzh.ch and andyj@ibb.waw.pl.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Andrzej Jerzmanowski (andyj@ibb.waw.pl).

A.J. and C.B. conceived the project; M.K., C.B., and A.J. designed the study; M.K. performed proteomic analyses; K.R. and M.L. performed transcriptomic analyses; L.K., A.M., M.K., and K.G. performed structural and phylogenetic in silico analyses and database screens; A.S. contributed the transcriptomic atlas of different tissues; M.K., C.B., and A.J. wrote the article.

[OPEN] Articles can be viewed without a subscription.
www.plantphysiol.org/cgi/doi/10.1104/pp.16.00214

neutralization upon DNA binding by its many Lys residues playing an important role (Hendzel et al., 2004). According to current models, simultaneous and synergistic binding of both GH1 and the CTD are prerequisites for correct H1 placement and determine its role in chromatin compaction (Stasevich et al., 2010). It is generally agreed that H1, by restricting nucleosome mobility and impeding the access of trans-acting factors to their target sequences, exerts strong effects on DNA-dependent activities, such as transcription and replication, and probably also recombination and repair (Izzo et al., 2008). Recent evidence suggests an even more complex pattern of H1 functions in the cell, in which its role as a universal architectural protein affecting chromatin dynamics is complemented by a parallel function as a local and gene-specific regulator (McBryant et al., 2010). Linker histones are a more divergent group of proteins than core histones. In animals, numerous nonallelic variants, including cell type- and stage-specific isoforms, have been described (Jerzmanowski, 2004; Sancho et al., 2008). In addition, and similar to core histones, major animal H1 variants undergo extensive posttranslational modifications of different types (Wisniewski et al., 2007), the importance of most of which is unknown.

Plant H1s exhibit the universal features of the H1 family, including the occurrence of different nonallelic variants and extensive posttranslational modifications (Table I; Supplemental Table S1; Prymakowska-Bosak et al., 1996; Jerzmanowski et al., 2000; Jerzmanowski, 2004; Kotliński et al., 2016). Interest in their functional roles has grown considerably in recent years, since they are frequently found in high-throughput screens aimed at identifying regulators involved in processes related to development, physiology, and adaptation to stresses (Wierzbicki and Jerzmanowski, 2005; She et al., 2013; Zemach et al., 2013; Over and Michaels, 2014; Rutowicz et al., 2015; Supplemental Table S2). However, because of the exceptional diversity of plant GH1-containing proteins, a fact not realized by most researchers, the relevant reference information about members of this group available in databases is highly imprecise, lacks coherence and systematization, and often is misleading, particularly for those unfamiliar with the classification of chromatin proteins. For example, as illustrated in Table I and Supplemental Table S1, plant linker histones, like high-mobility group A (HMGA) and certain other proteins, are described by the general term winged helix DNA-binding transcription factor in several databases. Numerous plant GH1-containing proteins are listed as putative or lack any description. Moreover, the annotation of the same proteins is inconsistent between databases.

Here, we summarize currently available information, including both published data and the findings of our in silico and high-throughput analyses, and propose a coherent system of phylogeny and structure-based nomenclature and annotation of H1s and other GH1-containing proteins of Arabidopsis (*Arabidopsis thaliana*). This system will be useful as a basic reference tool for the identification and characterization of homologous proteins from different plant species. In addition, we highlight some interesting trends in the evolution of chromatin-based regulation that may be specific for plants.

RESULTS AND DISCUSSION

The Arabidopsis genome encodes 15 proteins containing a genuine GH1 domain. A scheme linking GH1-based phylogenetic relationships with protein domain architectures within this group is shown in Figure 1. Phylogenetic analysis supports an early separation into three subgroups, which we rename here as follows: (1) H1s; (2) GH1-HMGA/GH1-HMGA-related; and (3) GH1-Myb/GH1-Myb-related. The above pattern is generally conserved in angiosperm plants, as shown by a maximum-likelihood phylogenetic tree of GH1-containing proteins from a broad range of plant species (Supplemental Fig. S1). The split into typical H1s and GH1-HMGA/GH1-HMGA-related preceded the separation of the GH1-Myb/GH1-Myb-related subgroup. The rapid diversification of the latter compared with the H1s suggests that it was not initially subjected to strong purifying selection but might have been important for the ongoing adaptive evolution of plants. Perhaps this could be the reason that genes encoding Arabidopsis GH1-containing proteins other than H1s show differential expression patterns in different tissues and developmental stages (Supplemental Fig. S3; Schmidt et al., 2011). Below, we discuss the properties of the three subgroups in more detail.

H1s

We have argued previously that the formal criteria that define a typical linker histone (i.e. a protein with a GH1 domain flanked by two unstructured and highly basic tails) are fulfilled by the products of only three Arabidopsis genes, designated *H1.1*, *H1.2*, and *H1.3* (Wierzbicki and Jerzmanowski, 2005). As shown in Figure 1, the subgroup of Arabidopsis H1s consists exclusively of this trio of H1s, none of which has any recognizable domain except GH1. Consistent with earlier analyses of phylogenetic relationships among known plant linker histones (Jerzmanowski et al., 2000; Rutowicz et al., 2015), this subgroup contains a representative (H1.3) of a distinct branch of stress-inducible H1 variants (Ascenzi and Gantt, 1997, 1999; Scippa et al., 2000, 2004; Przewloka et al., 2002; Jerzmanowski, 2007). Previously, we demonstrated that this branch separated from the main H1 variants roughly 140 million years ago, which coincided with the appearance of angiosperm plants on Earth (Rutowicz et al., 2015). There are no orthologs of stress-inducible H1 variants in sequenced species representing green algae, bryophytes, lycophytes, and conifers (gymnosperms; analyzed in Supplemental Fig. S1). Importantly, only members of the H1 subgroup

Table I. *Accession numbers and descriptions of Arabidopsis canonical linker histones in different databases: TAIR10, UniProt, NCBInr, and ChromDB (a copy of this discontinued database in the Web archive was used)*

Gene	Splice Variant	TAIR10		UniProt		NCBInr		ChromDB	Length
		Identifier	Description	Identifier	Description	Identifier	Description		
									amino acids
H1.1	1	AT1G06760.1	Winged-helix DNA-binding transcription factor family protein	P26568 (H11_ARATH)	Histone H1.1	NP_172161.1 P26568.1 AAF63139.1 AAL16244.1 CAA44314.1 AAK91467.1 AAM19868.1 AAM64441.1 AEE28032.1 CAA44312.1 (partial)	Histone H1.1, histone H1-1	HON1	274
H1.2	1	AT2G30620.1	Winged-helix DNA-binding transcription factor family protein	P26569 (H12_ARATH)	Histone H1.2	NP_180620.1 P26569.1 AAK25921.1 AAK64117.1 AEC08419.1 AAM63006.1 AAM15525.1 CAA44316.1	Histone H1.2, histone H1-2, putative histone H1 protein, histone H1	HON2	273
	2	AT2G30620.2	Winged-helix DNA-binding transcription factor family protein	F4INW2	Histone H1.2	NP_001189643.1 AEC08420.1	Histone H1.2		202
	3	–	–	C0Z3A1	AT2G30620 protein	BAH57180.1	AT2G30620		208
H1.3	1	AT2G18050.1	HIS1-3|histone H1-3	P94109	His-1-3 | histone H1	NP_179396.1 AAC49789.1 AAC49790.1 AAD20121.1 AAK76471.1 AAL85145.1 AAM61167.1 AEC06720.1	Histone H1-3, histone H1	HON3	167
	2	AT2G18050.2	HIS1-3|histone H1-3	Q3EBY3	Histone H1-3 HIS1-3	NP_849970.1 AEC06721.1	Histone H1-3		138

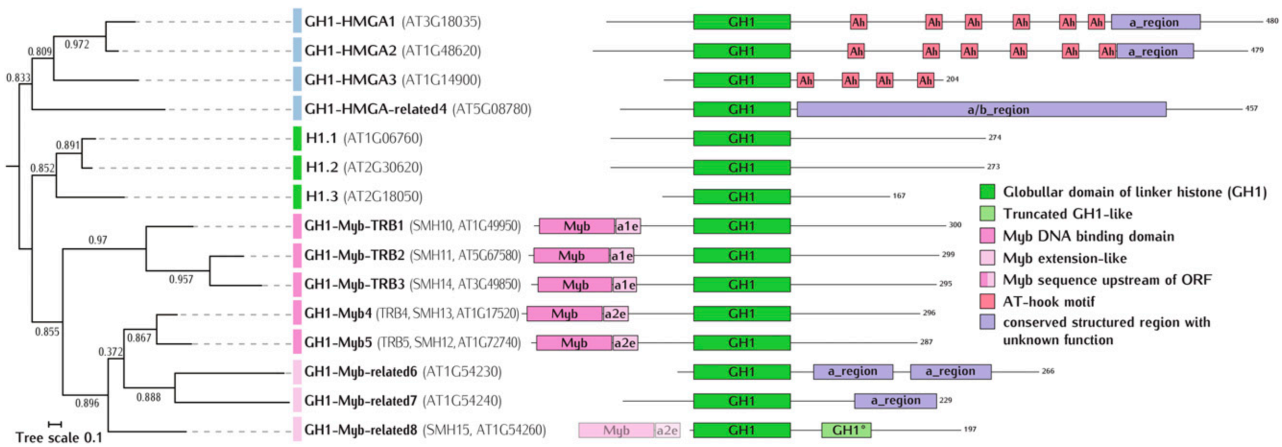


Figure 1. Maximum-likelihood phylogenetic tree and domain architecture of Arabidopsis GH1-containing proteins. Protein sequences were aligned with the local pair iterative algorithm implemented in Mafft (Yamada et al., 2016). Conserved columns from each multiple sequence alignment were selected manually. The phylogenetic analysis was performed with PhyML (Guindon et al., 2005), with the JTT model of amino acid substitutions and three random starting trees. Approximate likelihood ratio test SH-like (Shimodaira-Hasegawa-like) branch supports above 50% are shown. The tree was rooted using GH1-Myb as an internal sister outgroup for both GH1-HMGA and histone H1 clades. The tree image was prepared with iTol (Letunic and Bork, 2011). Domain architecture analysis was carried out using the SMART (Letunic et al., 2015) and GeneSilico (Kurowski and Bujnicki, 2003) Web servers and Meta-BASIC (Ginalski et al., 2004).

possess the characteristic regions of strong positive charge in all C-terminal and most N-terminal domains (Supplemental Fig. S2), beginning immediately adjacent to GH1. It should be noted that the regions of the N-terminal domains of H1.1 and H1.2 most distant from GH1 contain a negatively charged fragment that is targeted by posttranslational modification of phosphorylation, which further increases its negative charge (Kotliński et al., 2016). Thus, among Arabidopsis GH1-containing proteins, the pattern of charge distribution in the N- and C-terminal domains of H1s appears to be as distinctive a feature as the phylogenetic positions of their GH1s.

GH1-HMGA/GH1-HMGA-Related Versus Putative True Arabidopsis HMGA Proteins

In animals, HMGA proteins are distinguished by multiple AT-hook DNA-binding motifs: conserved nine-amino acid peptides capable of strong binding to 6-bp or longer AT-rich stretches of DNA via the minor groove. Except for an acidic C-terminal region, these proteins do not have any other recognized domains. In contrast, proteins currently defined in the literature as plant HMGA members contain a typical GH1 domain in addition to AT-hook motifs. This arrangement is restricted to angiosperm plants (Supplemental Fig. S1), suggesting a relatively late occurrence of GH1-AT-hook fusion in the evolution of plants. Arabidopsis has three such proteins (GH1-HMGA1 to GH1-HMGA3), which possess four to six AT-hook motifs. All three were detected in our analysis of the nuclear proteome of an Arabidopsis T87 cell suspension culture (Supplemental Table S1; http://proteome.arabidopsis.pl). Interestingly, the Arabidopsis GH1-HMGA cluster also includes a protein with no AT-hook domains (AT5G08780.1, named GH1-HMGA-related4 in our proposed nomenclature). We were unable to detect this protein in our T87 nuclear proteome (Supplemental Table S1), but its transcript was present in an Arabidopsis transcriptome derived by RNA sequencing analysis (Supplemental Table S1). Its GH1 sequence places GH1-HMGA-related4 distantly from the rest of the Arabidopsis H1-HMGA subgroup. Comparison of the charged amino acid profiles of non-GH1 fragments of Arabidopsis GH1-containing proteins demonstrated that the CTDs of GH1-HMGA1 to GH1-HMGA3 have an island-like distribution of positively and negatively charged residues, with mostly the latter present in fragments directly adjacent to GH1 (Supplemental Fig. S2). The corresponding profile for GH1-HMGA-related4 is significantly different. Secondary structure predictions suggest a potentially novel domain that lacks sequence similarity to any other protein domain of known or unknown structure/function. Interestingly, similar sequences are present in proteins from other species of the order Brassicales, in which they also are accompanied by GH1. The phylogenetic tree of GH1s from model plant proteomes identifies a distinct cluster composed of Arabidopsis GH1-HMGA-related4 and similar proteins from other species. Importantly, according to the InterPro database (http://www.ebi.ac.uk/interpro/), some of the proteins from other species belonging to this cluster retained AT-hook motifs.

The fusion of genuine GH1 and multiple AT-hook motifs that occurred in angiosperm plants also can be found in phylogenetic groups outside the plant kingdom, such as in numerous fish species, in *Trichoplax adhaerens*, the only extant representative of the phylum Placozoa (a primitive group of multicellular animals), as well as in some yeast, nematode, and insect species. The fish and *T. adhaerens* genomes encode very large proteins (up to 2,900 amino acids) in which GH1 and AT-hook motifs cooccur with RING and PHD domains. The other mentioned organisms possess simpler proteins in which GH1 coexists exclusively with AT-hook motifs. The phylogenetic relationships among these extremely diverse organisms suggest that multiple evolutionary events have resulted in the cooccurrence of GH1 and AT-hook motifs within their proteins.

Surprisingly given the fundamental functions of HMGA proteins in animals, the functional significance of the GH1/multiple AT-hook motif fusion has never been studied, despite its being referred to in all the major literature concerning plant HMG proteins. Notably, in several prokaryotes in which either HMGA-like or histone H1 CTD-like domains are present in important hub proteins regulating critical cellular processes, these two domains were found to be functionally equivalent and could be interchanged without any phenotypic consequences. Moreover, even chimeras in which the AT-hook domain was substituted by the human histone H1 CTD or full-length human H1 functioned properly in prokaryotic hosts (García-Heras et al., 2009). Thus, Arabidopsis GH1-HMGA proteins may be considered as highly specialized derivatives of H1 in which the typical CTD of H1 has been replaced by HMGA. To try and verify such a possibility, we reexamined the long-held view that Arabidopsis is devoid of canonical HMGA proteins. Using the SMART tool (Schultz et al., 2000; http://smart.embl-heidelberg.de), we identified 48 Arabidopsis proteins containing AT-hook motifs, 23 of which, unlike typical HMGA members, contain only a single AT hook. Most of the identified proteins, including those of the H1-HMGA subgroup, contain additional domains. Only two proteins, the predicted products of the alternatively spliced *At1g48610* gene, contain four AT-hook motifs and no other domain. *At1g48610.1* encodes a relatively small protein (212 amino acids, about 21.6 kD) with a high pI (pI = 11.6), features typical for HMGA. *At1g48610.2* (transcript retains the last intron) encodes a shorter protein with a pI of 11.4. The other putative proteins with the AT-hook motif are significantly larger, and their pI, unlike that of canonical HMGA, is below 10. Interestingly, a protein encoded by *At1g48610* was detected in our analyses of the nuclear proteome of Arabidopsis T87 cells, with a score and peptide number similar to those of core and linker histones, which indicated a substantial concentration in nuclei (http://proteome.arabidopsis.pl). Moreover, and probably due to its high pI, it was copurified during the isolation of Arabidopsis linker histones by extraction with 4.5% PCA (perchloric acid) and cation-exchange chromatography (Kotliński et al., 2016).

In both analyses, the larger version of AT1G48610 had a higher number of peptides and a higher score than the smaller form (100% and 92% of sequence coverage, respectively). Using four different proteases (trypsin, ArgC, termolysin, and pepsin), we identified 516 peptides unique for AT1G48610.1 (i.e. matching the last 29 amino acids of this protein), including peptides spanning the exon-exon junction. However, we detected no peptides unique for the smaller AT1G48610.2 form (i.e. matching the last 14 amino acids that are different in this variant). Similarly, RNA sequencing analysis revealed multiple reads spanning the junction of the last two exons of the gene but only one low-quality read within the intron retained in AT1G48610.2. These data indicate that the larger version of the protein (AT1G48610.1) is the main product of this gene. According to the BAR Toronto database (Toufighi et al., 2005), the expression of *At1g48610* is strongest in the central, rib, and peripheral zones of the shoot apical meristem, in pistil tissue primarily consisting of ovaries, and in phloem companion cells at the border of the meristematic and elongation zones of the root. This suggests that AT1G48610, which we believe to be a true Arabidopsis HMGA protein, is important in the differentiation of stem cells, a role highly reminiscent of that played by animal HMGA-type proteins (Ozturk et al., 2014). Interestingly, the *At1g48610* locus in chromosome 1 is located next to that encoding the H1-HMGA2 protein.

GH1-Myb/GH1-Myb-Related

This subgroup comprises five proteins with an additional N-terminal Myb domain accompanied by a 17- to 18-amino acid-long Myb extension-like domain. They seem to be as evolutionarily old as H1s, as, in addition to angiosperms, they occur in representatives of green algae, bryophytes, lycophytes, and gymnosperms (Supplemental Fig. S1). They are known as Single Myb Histone (SMH) or Telomere Repeat Binding (TRB) proteins, and two of them, GH1-Myb-TRB1 and GH1-Myb-TRB2, were shown to bind Arabidopsis telomeric repeats in vitro through a Myb domain of the telobox (telomere motif AAACCCTAA) type (Marian et al., 2003; Schrumpfová et al., 2004). The demonstration of in vivo interactions of these proteins with Arabidopsis telomerase supports a suggestion that they are part of the greater plant telomeric interactome (Schrumpfová et al., 2014). However, a recent mapping by chromatin immunoprecipitation sequencing of the genome-wide distribution of TRB1:GFP revealed its presence in over 7,800 genomic loci. The majority of these loci contained telobox-related motifs located at the transcription start sites, with additional loci spreading across gene bodies as well as distal promoter regions. Moreover, it was shown by genome-wide expression (RNA sequencing) analysis that TRB1, by binding at these loci, plays the role of transcriptional regulator, which is independent of its role in telomere maintenance (Zhou et al., 2016). Given such widespread occurrence, it seems highly probable that, at least in some of the detected loci, TRB1, through its GH1 domain, competes for nucleosome binding with H1s.

Since GH1-Myb-TRB3 is very similar to GH1-Myb-TRB1 and GH1-Myb-TRB2 (all three locate on the same branch of the phylogenetic tree; Supplemental Fig. S1), it may perform the same function. GH1-Myb-TRB1 was identified in our proteomic analysis of Arabidopsis nuclei, while GH1-Myb-TRB2 and GH1-Myb-TRB3 were detected below the established threshold (Supplemental Table S1). Transcripts encoding GH1-Myb-TRB1 to GH1-Myb-TRB3 were all present in our RNA sequencing data. Two other GH1-Myb proteins, GH1-Myb4 and GH1-Myb5 (AT1G17520.1 and AT1G72740.1, respectively), are more distantly related to GH1-Myb-TRB1 to GH1-Myb-TRB3 (Supplemental Fig. S1). The three other proteins of this subgroup (GH1-Myb-related6 to GH1-Myb-related8) lack the Myb domain, although the transcript of one them (AT1G54260.1) contains a Myb-coding sequence in front of the start codon, suggesting the loss of this domain during evolution. AT1G54260.1 also contains a strongly diverged and truncated GH1 domain at the C-terminal side of its regular GH1 domain. According to secondary structure predictions, the two other proteins lacking the Myb domain (AT1G54230 and AT1G54240) have α-helical regions within their CTDs. Interestingly, all three proteins lacking Myb are encoded by neighboring genes on chromosome 1. The N- and C-terminal domains of all proteins from the GH1-Myb/GH1-Myb-related subgroup are mostly negatively charged.

A Rationale for the Proposed New Nomenclature of Arabidopsis GH1-Containing Proteins

At first glance, the evolutionary diversification of H1s into well-distinguished and conserved subtypes seems to be less pronounced in angiosperm plants than in animals, particularly vertebrates. The most distinct structural and functional diversification of plant H1s coincided with the appearance of angiosperms (approximately 140 million years ago) and resulted in two major subtypes that have been maintained ever since: the main and stress-inducible H1s. Regarding H1s, the case of Arabidopsis shows that two main variants and a single stress-inducible variant are sufficient to support the basic processes of growth and development in a typical flowering plant. While this does not rule out the functional significance of more subtle variation within these two major subtypes observed in systematically distant families and species, proof of such significance has yet to be provided. The above notwithstanding, the impression of a seemingly limited diversification of H1s during the evolution of plants may be misleading and result from biased classification rules. These rules were adopted from studies on typical animal H1s and do not take into account the fundamentally different life strategies and vastly different selection pressures shaping major chromatin structural proteins in plants and animals during their long histories of separate evolution. The GH1-HMGA/GH1-HMGA-related and GH1-Myb/GH1-Myb-related subgroups could be the end result of such specific selection pressures in the plant kingdom. The concept that proteins of these two subgroups represent highly diverged and specialized derivatives of plant H1 that use GH1 as a common motif for targeting nucleosomes is supported by the conserved phylogenetic relationships among plant GH1-containing proteins, a recently demonstrated widespread occurrence of GH1-Myb-TRB1 in chromatin, and its likely involvement in transcriptional regulation, as well as by the identification of a candidate for a true Arabidopsis HMGA protein that does not contain a GH1 domain. This concept is by no means equivalent to suggesting that all plant GH1-containing proteins are bona fide H1 variants, in a sense ascribed to this subcategory in animal studies. Its main purpose is to draw attention to the fact that, in plants, the competition-based removal of H1 from chromatin may be dependent on a more diversified and specialized group of competitors than in animals, suggesting novel plant-specific mechanisms of chromatin regulation.

Therefore, we propose a unified nomenclature for plant GH1-containing proteins built simply on their GH1-based phylogenetic relationships, as shown in Figure 1. We further propose to distinguish proteins possessing two characteristic domains (GH1-HMGA and GH1-Myb) and proteins belonging to the same subgroups due to the phylogenetic position of their GH1 but lacking the second characteristic domain, HMGA or Myb. We name these latter proteins GH1-HMGA-related and GH1-Myb-related, respectively (they are marked by lighter color in Supplemental Fig. S1). It is important to remember that proteins of these two types from other species still retain their AT-hook motifs and Myb domains. Since the GH1-Myb-TRB1 and GH1-Myb-TRB2 proteins have been experimentally confirmed to bind telomere repeats and, therefore, were named TRB1 and TRB2, we propose to retain this functional reference in their names (as GH1-Myb-TRB) for the sake of clarity and tradition. The same applies to GH1-Myb-TRB3, a very similar protein that has been described previously as TRB3. With regard to GH1-Myb4 and GH1-Myb5, which also are described as TRB proteins in many databases, we suggest removing the designation TRB from their names. In the Arabidopsis GH1 evolutionary tree, both of these proteins group in a clade separate from that of TRB1 to TRB3, suggesting a greater evolutionary distance. Moreover, and unlike GH1-Myb-TRB1 to GH1-Myb-TRB3, they both contain a Myb extension-like sequence different from GH1-Myb-TRB1 to GH1-Myb-TRB3, so their binding preferences may be different. We also have indicated (Supplemental Table S1; Supplemental Fig. S1, parentheses) the former names of GH1-Myb proteins as SMH that were used in the discontinued ChromDB and in maize (*Zea mays*) genomic databases. Importantly, our inspection in SMART/UniProt of the domain structures of all proteins included in the tree in Supplemental Figure S1 revealed some singularities. In *Medicago truncatula*, a GH1-Myb protein has an additional RNA-recognition motif. Another GH1-Myb of this species has a strongly changed GH1 domain. Both *Brassica rapa* and *Oryza sativa* have a GH1-Myb protein carrying an additional domain, and maize contains a GH1-HMGA protein with an S/T kinase domain. Moreover, in the maize H1 group, there is a protein with two AT-hook motifs (indicative that such fusions are not unusual in plants). While exception proves the rule, it cannot be excluded that at least some of the above singularities resulted from errors in genome assemblies or gene models.

We believe that the proposed phylogeny- and structure-supported system of classification, apart from practical convenience, will foster novel approaches in studies on the functional roles of GH1-containing proteins in plants.

MATERIALS AND METHODS

Database Screen

All proteins from TAIR (http://arabidopsis.org) and protein records from Arabidopsis (*Arabidopsis thaliana*) deposited in the NCBInr (https://www.ncbi.nlm.nih.gov/) and UniProt (http://www.uniprot.org/) databases were searched with the use of BLAST (Altschul et al., 1990) for proteins containing a GH1 domain. Sequences of GH1 from all 15 Arabidopsis GH1-containing proteins were used as queries. All records found are included in Table I and Supplemental Table S1 (Fucile et al. 2011). Additionally, the full genomic sequence from TAIR repository was translated in six reading frames and searched by position-specific iterated BLAST (Altschul et al., 1997). All 15 GH1 sequences from known proteins were used as queries. We have not found any new GH1-containing proteins in Arabidopsis.

Domain Architecture

Domain architecture analysis was carried out for all Arabidopsis proteins containing a GH1 domain using Meta-BASIC (Ginalski et al., 2004) as well as SMART (Letunic et al., 2015) and GeneSilico (Kurowski and Bujnicki, 2003) Web servers. The regions with no detectable homology to known protein domains, yet with conserved sequence and predicted secondary structures (with PSIPRED; Jones, 1999), also have been denoted as potential new domains.

For proteins assigned previously to the GH1-Myb subfamily yet lacking the Myb domain, nucleotide upstream/downstream sequences of coded genes were verified using both manual translations and data from TAIR gene model and exon confidence ranking system (https://www.arabidopsis.org/download_files/Genes/TAIR10_genome_release/TAIR10_gene_confidence_ranking/DOCUMENTATION_TAIR_Gene_Confidence.pdf). The truncated GH1 domain detected in AT1G54260 was verified in a similar manner.

Moving-Sum Plot

A moving-sum plot of net charge was generated for both N- and C-terminal regions (with respect to the GH1 domain) of all Arabidopsis GH1-containing proteins. The net charge was summed in a 20-amino acid sliding window along N- and C-terminal regions, starting from the GH1 domain. For each region, the percentages of both positively (K, R) and negatively (D, E) charged residues, total charge, and theoretical pI (calculated with http://web.expasy.org/compute_pi) also were calculated.

Phylogenetic Analyses

Protein sequences for model plants were collected via phmmer (Finn et al., 2011), available from the Ensembl Plants Web site (Kersey et al., 2014). The Ensembl database was chosen to ensure data quality, limiting the data set to well-studied organisms with possibly complete proteomes. This data set enables observations of specific subfamily expansions (due to consecutive duplications) in some angiosperms from Brassicaceae and Fabaceae. For better taxon sampling, the following representatives of missing major taxon groups were added: *Auxenochlorella protothecoides*, *Coccomyxa subellipsoidea*, *Marchantia polymorpha*, *Picea sitchensis*, *Pinus taeda* (from UniProt), and *Klebsormidium flaccidum* (from NCBI genomes).

Sequence searches were performed using H1.2, GH1-HMGA2, GH1-Myb-TRB1, and TRB1 from Arabidopsis as queries. All hits were mapped on UniProt identifiers (http://www.uniprot.org), except for *Physcomitrella patens* (which lacks UniProt identifiers for two out of nine analyzed sequences). Subsequently, representative plants were chosen with emphasis on Brassicaceae (three taxa) and including all basal plant model organisms present in the aforementioned database (for a list of identifiers and names, see Supplemental Table S3). Incomplete truncated sequences were discarded. Phylogenetic trees were inferred both for Arabidopsis GH1 proteins (Fig. 1) and for 282 representative plant sequences (Supplemental Fig. S1).

Sequences of all GH1-containing proteins used for phylogenetic comparison were screened with SMART (Schultz et al., 2000; Letunic et al., 2015) for the presence of any additional domains or loss of domains (other than GH1). The results are included in Supplemental Figure S1.

Accession Numbers

Sequence data from this article can be found are provided in tables, figures and Supplemental Data.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Maximum-likelihood phylogenetic tree of GH1-containing proteins from selected plants.

Supplemental Figure S2. Moving-sum plot of net charge for N- and C-terminal domains of all Arabidopsis GH1-containing proteins.

Supplemental Figure S3. Relative expression levels of GH1-containing protein-coding genes in Arabidopsis across 74 tissue- or cell-specific microarrays.

Supplemental Table S1. Accession numbers and descriptions of Arabidopsis proteins containing a GH1 domain from different databases (TAIR10, UniProt, NCBInr, and ChromDB).

Supplemental Table S2. List of articles referring to the role of plant linker histones.

Supplemental Table S3. List of GH1-containing protein identifiers in selected model plants.

Supplemental Methods. Supplemental materials and methods.

ACKNOWLEDGMENTS

We thank Fred Berger, Franziska Turck, Fredy Barneche, J. Mark Cock, Lars Hennig, Motoaki Seki, Anna Amtmann, and Doris Wagner for comments and fruitful discussions.

Received February 13, 2017; accepted March 10, 2017; published March 15, 2017.

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403–410

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Ascenzi R, Gantt JS (1997) A drought-stress-inducible histone gene in Arabidopsis thaliana is a member of a distinct class of plant linker histone variants. Plant Mol Biol **34:** 629–641

Ascenzi R, Gantt JS (1999) Molecular genetic analysis of the drought-inducible linker histone variant in Arabidopsis thaliana. Plant Mol Biol **41:** 159–169

Brown DT, Izard T, Misteli T (2006) Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. Nat Struct Mol Biol **13:** 250–255

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res **39:** W29–W37

Fucile G, Di Biase D, Nahal H, La G, Khodabandeh S, Chen Y, Easley K, Christendat D, Kelley L, Provart NJ (2011) ePlant and the 3D data display initiative: integrative systems biology on the world wide web. PLoS ONE **6:** e15237

García-Heras F, Padmanabhan S, Murillo FJ, Elías-Arnanz M (2009) Functional equivalence of HMGA- and histone H1-like domains in a bacterial transcriptional factor. Proc Natl Acad Sci USA **106:** 13546–13551

Ginalski K, von Grotthuss M, Grishin NV, Rychlewski L (2004) Detecting distant homology with Meta-BASIC. Nucleic Acids Res **32:** W576–W581

Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online: a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res **33:** W557–W559

Hansen JC, Lu X, Ross ED, Woody RW (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains. J Biol Chem **281:** 1853–1856

Hendzel MJ, Lever MA, Crawford E, Th'ng JPH (2004) The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo. J Biol Chem **279:** 20028–20034

Izzo A, Kamieniarz K, Schneider R (2008) The histone H1 family: specific members, specific functions? Biol Chem **389:** 333–343

Jerzmanowski A (2004) The linker histones. *In* Chromatin Structure and Dynamics: State-of-the-Art. New Comprehensive Biochemistry, **Vol 39**. Elsevier, Amsterdam, The Netherlands pp 75–102

Jerzmanowski A (2007) SWI/SNF chromatin remodeling and linker histones in plants. Biochim Biophys Acta **1769:** 330–345

Jerzmanowski A, Przewłoka M, Grasser KD (2000) Linker histones and HMG1 proteins of higher plants. Plant Biol **2:** 586–597

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292:** 195–202

Kersey PJ, Allen JE, Christensen M, Davis P, Falin LJ, Grabmueller C, Hughes DST, Humphrey J, Kerhornou A, Khobova J, et al (2014) Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res **42:** D546–D552

Kotliński M, Rutowicz K, Kniżewski Ł, Palusiński A, Olędzki J, Fogtman A, Rubel T, Koblowska M, Dadlez M, Ginalski K, et al (2016) Histone H1 variants in Arabidopsis are subject to numerous post-translational modifications, both conserved and previously unknown in histones, suggesting complex functions of H1 in plants. PLoS ONE **11:** e0147908

Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Res **31:** 3305–3307

Letunic I, Bork P (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res **39:** W475–W478

Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status in 2015. Nucleic Acids Res **43:** D257–D260

Marian CO, Bordoli SJ, Goltz M, Santarella RA, Jackson LP, Danilevskaya O, Beckstette M, Meeley R, Bass HW (2003) The maize Single myb histone 1 gene, Smh1, belongs to a novel gene family and encodes a protein that binds telomere DNA repeats in vitro. Plant Physiol **133:** 1336–1350

McBryant SJ, Lu X, Hansen JC (2010) Multifunctionality of the linker histones: an emerging role for protein-protein interactions. Cell Res **20:** 519–528

Over RS, Michaels SD (2014) Open and closed: the roles of linker histones in plants and animals. Mol Plant **7:** 481–491

Ozturk N, Singh I, Mehta A, Braun T, Barreto G (2014) HMGA proteins as modulators of chromatin structure during transcriptional activation. Front Cell Dev Biol **2:** 5

Prymakowska-Bosak M, Przewłoka MR, Iwkiewicz J, Egierszdorff S, Kuraś M, Chaubet N, Gigot C, Spiker S, Jerzmanowski A (1996) Histone H1 overexpressed to high level in tobacco affects certain developmental programs but has limited effect on basal cellular functions. Proc Natl Acad Sci USA **93:** 10250–10255

Przewloka MR, Wierzbicki AT, Slusarczyk J, Kuraś M, Grasser KD, Stemmer C, Jerzmanowski A (2002) The "drought-inducible" histone H1s of tobacco play no role in male sterility linked to alterations in H1 variants. Planta **215:** 371–379

Rutowicz K, Puzio M, Halibart-Puzio J, Lirski M, Kotliński M, Kroteń MA, Knizewski L, Lange B, Muszewska A, Śniegowska-Świerk K, et al (2015) A specialized histone H1 variant is required for adaptive responses to complex abiotic stress and related DNA methylation in Arabidopsis. Plant Physiol **169:** 2080–2101

Sancho M, Diani E, Beato M, Jordan A (2008) Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. PLoS Genet **4:** e1000227

Schmidt A, Wuest SE, Vijverberg K, Baroux C, Kleen D, Grossniklaus U (2011) Transcriptome analysis of the Arabidopsis megaspore mother cell uncovers the importance of RNA helicases for plant germline development. PLoS Biol **9:** e1001155

Schrumpfová P, Kuchar M, Miková G, Skrísovská L, Kubicárová T, Fajkus J (2004) Characterization of two Arabidopsis thaliana myb-like proteins showing affinity to telomeric DNA sequence. Genome **47:** 316–324

Schrumpfová PP, Vychodilová I, Dvořáčková M, Majerská J, Dokládal L, Schořová S, Fajkus J (2014) Telomere repeat binding proteins are functional components of Arabidopsis telomeres and interact with telomerase. Plant J **77:** 770–781

Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res **28:** 231–234

Scippa GS, Di Michele M, Onelli E, Patrignani G, Chiatante D, Bray EA (2004) The histone-like protein H1-S and the response of tomato leaves to water deficit. J Exp Bot **55:** 99–109

Scippa GS, Griffiths A, Chiatante D, Bray EA (2000) The H1 histone variant of tomato, H1-S, is targeted to the nucleus and accumulates in chromatin in response to water-deficit stress. Planta **211:** 173–181

She W, Grimanelli D, Rutowicz K, Whitehead MWJ, Puzio M, Kotlinski M, Jerzmanowski A, Baroux C (2013) Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. Development **140:** 4008–4019

Stasevich TJ, Mueller F, Brown DT, McNally JG (2010) Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. EMBO J **29:** 1225–1234

Syed SH, Goutte-Gattat D, Becker N, Meyer S, Shukla MS, Hayes JJ, Everaers R, Angelov D, Bednar J, Dimitrov S (2010) Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. Proc Natl Acad Sci USA **107:** 9620–9625

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The Botany Array Resource: e-northerns, expression angling, and promoter analyses. Plant J **43:** 153–163

Wierzbicki AT, Jerzmanowski A (2005) Suppression of histone H1 genes in Arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation. Genetics **169:** 997–1008

Wisniewski JR, Zougman A, Kruger S, Mann M (2007) Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. Mol Cell Proteomics **6:** 72–87

Woodcock CL, Skoultchi AI, Fan Y (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome Res **14:** 17–25

Yamada KD, Tomii K, Katoh K (2016) Application of the MAFFT sequence alignment program to large data: reexamination of the usefulness of chained guide trees. Bioinformatics **32:** 3246–3251

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell **153:** 193–205

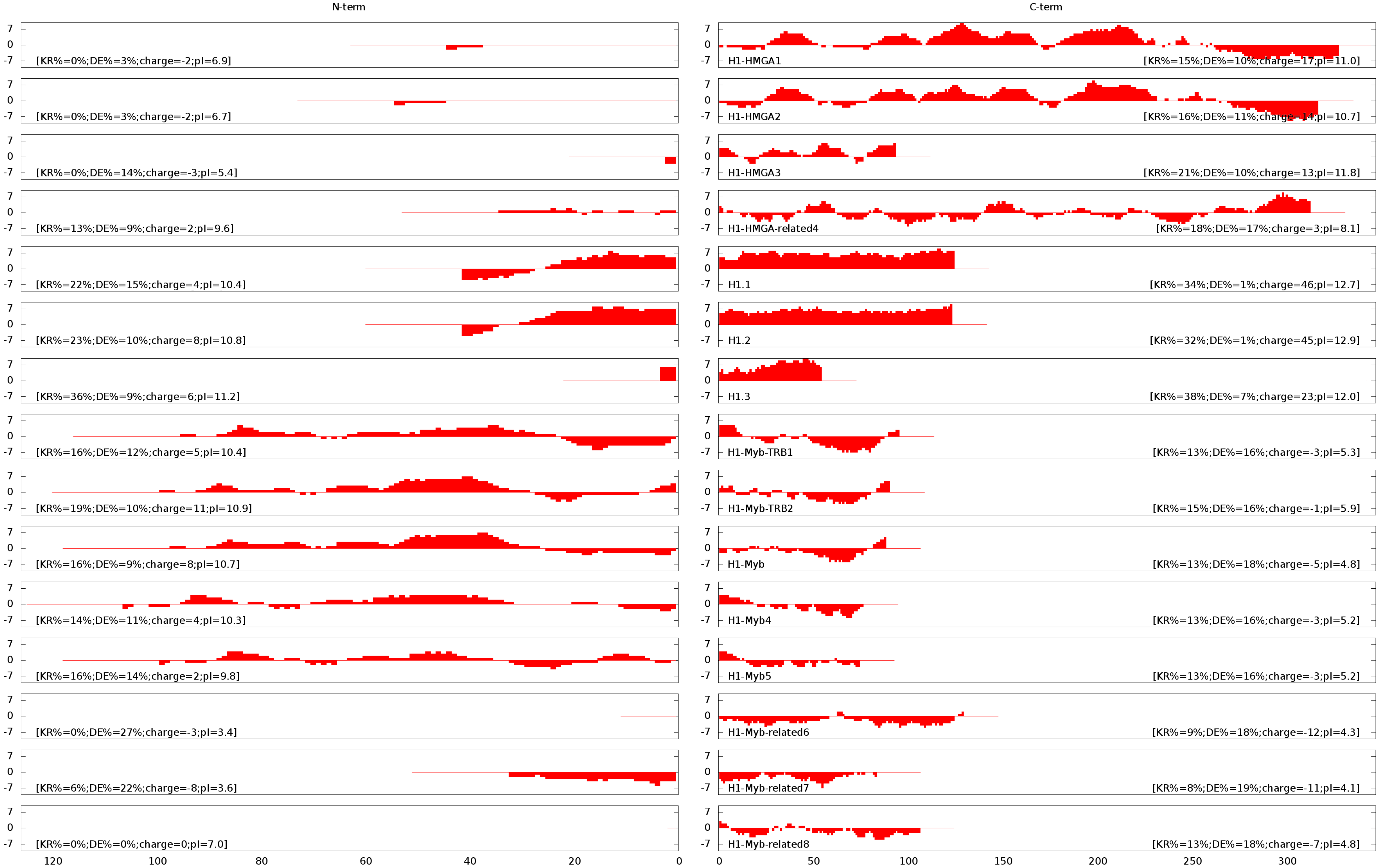
Zhou BR, Feng H, Kato H, Dai L, Yang Y, Zhou Y, Bai Y (2013) Structural insights into the histone H1-nucleosome complex. Proc Natl Acad Sci USA **110:** 19390–19395

Zhou Y, Hartwig B, James GV, Schneeberger K, Turck F (2016) Complementary activities of TELOMERE REPEAT BINDING proteins and polycomb group complexes in transcriptional regulation of target genes. Plant Cell **28:** 87–101

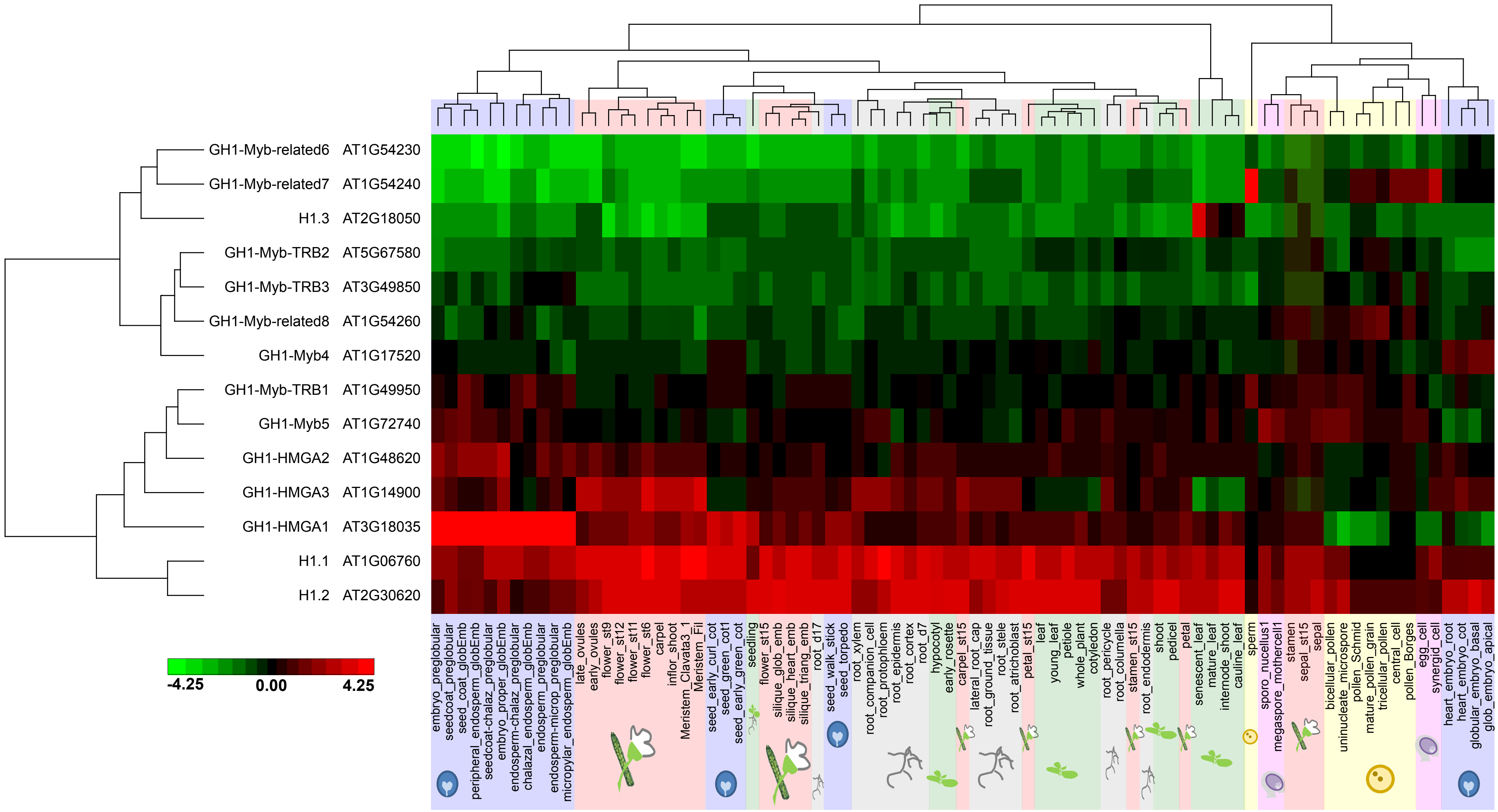
Zhou YB, Gerchman SE, Ramakrishnan V, Travers A, Muyldermans S (1998) Position and orientation of the globular domain of linker histone H5 on the nucleosome. Nature **395:** 402–405

Supplementary Fig. S1 Maximum-likelihood phylogenetic tree of GH1-containing proteins from selected plants.

Suplementary Fig. S2. Moving sum plot of net charge for N- and C- terminal domains of all Arabidopsis GH1–containing proteins. The net charge (y-axis) is summed in a 20–aa sliding window, with the position along the N– and C–terminal domains, with respect to GH1, denoted on the x-axis. For each N- and C-terminus, the percentages of both positively (K, R) and negatively (D, E) charged residues, total charge and theoretical isoelectric point (pI, calculated with http://web.expasy.org/compute_pi) are also shown.



Supplementary Fig. S3. Relative expression levels of GH1-containing protein coding genes in Arabidopsis, across 74 tissue– or cell–specific microarrays (as used in Schmidt et al., 2011). The arrangement of genes and samples is based on euclidean distance and hierachical agglomerative clustering. Colors are scaled per row. Red and green ranges correspond to high and low expression levels, respectively. The pictograms indicate the cell and tissue types.



Supplementary Table S1. Accession numbers and descriptions of Arabidopsis thaliana proteins containing a GH1 domain from different databases (TAIR10, UniProt, NCBI and ChromDB*). This table is supplemented with data concerning the occurrence, rank and scores of proteins identified in the nuclear proteome of Arabidopsis T-87 suspension culture cells (6753 proteins identified in total, www.proteome.arabidopsis.pl) and localization of gene expression according to the BAR Toronto database (Fucile et al., 2011) (http://bar.utoronto.ca/). *a copy of this discontinued database in the web archive was used.

	Symbol	TAIR		Uniprot		len gth	NCBI		ChromDB	Nuclear protemome			BAR Toronto
		AGI / *locus*	Description	ID	Description		ID	Description		rank	score	presence	
1	**GH1-HMGA1**	AT3G18035.1	HON4 | winged-helix DNA-binding transcription factor family protein	Q9LSK7	HON4 At3g18035	480	ref|NP_188431.3 dbj|BAB01332.1 gb|AAO00794.1 gb|AAP31950.1 gb|AEE76037.1	HON4; unnamed protein product; linker histone protein, putative; At3g18035; winged-helix DNA-binding transcription factor family protein	-	1458	843	+	most tissues, not pollen
2	**GH1-HMGA2**	AT1G48620.1	HON5 | high mobility group A5	Q4V3D1	HON5 At1g48620	479	ref|NP_175295.1 gb|AAY56416.1 gb|ABF57274.1 gb|AEE32328.1	At1g48620; high mobility group A5	HMGA2	1017	963	+	apical meristem, most other tissues
3		*AT1G48620*		Q9LP61	T1N15.25	594	gb|AAF79708.1|AC020889_16	T1N15.25				+	
4		*AT1G48620*		Q9C6X5	Putative uncharacterized protein F9P7.3	332	gb|AAG50847.1|AC074308_3	hypothetical protein, 3' partial				+	
5	**GH1-HMGA3**	AT1G14900.1	HMGA | high mobility group A	Q43386	HMG-Y-related protein A | At1g14900,F10B6.31	204	ref|NP_172943.1 sp|Q43386.1|HMGYA_ARATH gb|AAF79232.1|AC006917_17 emb|CAA67564.1 gb|AAB97739.1 gb|AAO44072.1 dbj|BAH19921.1 gb|AEE29240.1 emb|CAA71797.1 (one mismatch)	high mobility group protein A; F10B6.31; HMG-Y-related protein A; HMG-I/Y protein	HMGA3	3144	122	+	apical meristem, most other tissues, phloem companion cells
6	**GH1-HMGA-related4**	AT5G08780.1	winged-helix DNA-binding transcription factor family protein	Q6AWW7	At5g08780	457	ref|NP_680160.2 gb|AAT85727.1 gb|AAU94419.1 gb|AED91349.1	winged-helix DNA-binding transcription factor family protein; At5g08780	HMGA4	-	-	-	no data
7		*AT5G08780*		Q9C599	Putative uncharacterized protein At5g08780	463	emb|CAC35883.1	putative protein					
8	**GH1-Myb-TRB1** (SMH10)	AT1G49950.1	TRB1, ATTRB1 | telomere repeat binding factor 1	Q8VWK4	Telomere repeat-binding factor 1 | TRB1 At1g49950,F2J10.16 (Identical sequences of proteins in all splice variants)	300	ref|NP_564559.1 ref|NP_849789.1 ref|NP_973998.1 sp|Q8VWK4.1|TRB1_ARATH gb|AAL73123.1|U83623_1 gb|AAL32814.1 gb|AAP80178.1 gb|AAS10009.1 gb|AEE32497.1 gb|AEE32498.1 gb|AEE32499.1 gb|AAL73438.1|U83624_1 (one mismatch) gb|AAM65540.1 (two mismatches)	telomere repeat binding factor 1; MYB transcription factor; Unknown protein; At1g49950; DNA-binding protein PcMYB1, putative	SHM10	688	3025	+	mature pollen, most other tissues
9		AT1G49950.2	TRB1, ATTRB1 | telomere repeat binding factor 1										
10		AT1G49950.3	TRB1, ATTRB1 | telomere repeat binding factor 1										
11		*AT1G49950*				318	gb|AAF76448.1|AC015445_15	Contains similarity to DNA-binding protein MYB1 from Petroselinum crispum gi|7488946 and contains MYB-DNA-binding PF|00249 and linker-Histone PF|00538 domains					

12	**GH1-Myb-TRB2** (SMH11)	AT5G67580.1	TRB2, ATTRB2, TBP3, ATTBP3 | Homeodomain-like/winged-helix DNA-binding family protein	Q9FJW5	Telomere repeat-binding factor 2 | TRB2 (Identical sequences of proteins in both splice variants)	299	ref|NP_201559.1 ref|NP_851286.1 sp|Q9FJW5.1|TRB2_ARATH gb|AAL73442.1|U83836_1 dbj|BAB08466.1 gb|AAK63987.1 gb|AAL76146.1 gb|AAS10015.1 gb|AED98362.1 gb|AED98363.1 gb|AAL73441.1|U83837_1 (one mismatch)	telomere repeat binding factor 2; MYB transcription factor; Telomere-binding protein 3; unnamed protein product; AT5g67580/K9I9_15;	SMH11	4281	181	bellow treshold, detected in 2 samples only	most tissues
13		AT5G67580.2	TRB2, ATTRB2, TBP3, ATTBP3 | Homeodomain-like/winged-helix DNA-binding family protein										
14		*AT5G67580*		G0XQD5	Truncated telomeric DNA binding protein isoform	190	gb|AEK67481.1	truncated telomeric DNA binding protein isoform					
15	**GH1-Myb-TRB3** (SMH14)	AT3G49850.1	TRB3, ATTRB3, TBP2 | telomere repeat binding factor 3	Q9M2X3	Telomere repeat-binding factor 3 | TRB3 TBP2,At3g49850,T16K5.200	295	ref|NP_190554.1 sp|Q9M2X3.1|TRB3_ARATH gb|AAL73439.1|U83839_1 gb|AAL73440.1|U83838_1 emb|CAB66923.1| gb|AAL24273.1| gb|AAL57702.1| gb|AAL79593.1| gb|AAS10012.1| gb|AEE78598.1|	telomere repeat binding factor 3; MYB transcription factor; Telomere-binding protein 2; MYB-like protein; AT3g49850/T16K5_200	SMH14	5906	31	bellow treshold, only one peptide detected, low score	most tissues
16	**GH1-Myb4** (TRB4, SMH13)	AT1G17520.1	Homeodomain-like/winged-helix DNA-binding family protein	F4I7L1	Telomere repeat-binding factor 4 | At1g17520,F1L3.23	296	sp|F4I7L1.2|TRB4 dbj|BAC43136.1 gb|AAO63354.1 gb|AAS10008.1 ref|NP_173195.2| (one mismatch) gb|AEE29601.1| (one mismatch)	Telomere repeat-binding factor 4; putative telomere repeat-binding factor 4; MYB transcription factor; putative myb-related DNA-binding protein; At1g1752	SMH13	-	-	-	dry seed
17		*AT1G17520*				240	gb|AAF79481.1|AC022492_25	F1L3.23					
18	**GH1-Myb5** (TRB5, SMH12)	AT1G72740.1	Homeodomain-like/winged-helix DNA-binding family protein			289	gb|AAG51858.1|AC010926_21	putative DNA-binding protein; 27830-29933	SMH12	5374	57	bellow treshold, only one peptide detected	mature pollen, most other tissues
19		*AT1G72740*		F4IEY4	Telomere repeat-binding factor 5 | At1g72740,F28P22.7	287	ref|NP_001077814.1 gb|AEE35367.1	homeodomain-like/winged-helix DNA-binding protein					
20		AT1G72740.2	Homeodomain-like/winged-helix DNA-binding family protein	F4IEY3	Homeodomain-like/winged-helix DNA-binding protein	281	ref|NP_177418.2 sp|F4IEY4.1|TRB5_ARATH gb|AEE35366.1	homeodomain-like/winged-helix DNA-binding protein; Telomere repeat-binding factor 5; MYB transcription factor; homeodomain-like/winged-helix DNA-binding protein					
21		*AT1G72740*				151	gb|AAK50065.1|AF372925_1 gb|AAM70558.1	At1g72740/F28P22_7					
22	**GH1-Myb-related6**	AT1G54230.1	Winged helix-turn-helix transcription repressor DNA-binding	F4HV91	Winged helix-turn-helix transcription repressor DNA-binding protein	232	ref|NP_175825.2 gb|AEE33069.1	winged helix-turn-helix transcription repressor DNA-binding protein	-	-	-	-	flower buds, mature pollen, cotylodones
23		*AT1G54230*		Q9SLK9	Putative uncharacterized protein F20D21.5	276	gb|AAD25603.1|AC005287_5	Hypothetical protein					

24	**GH1-Myb-related7**	AT1G54240.1	winged-helix DNA-binding transcription factor family protein	Q1PFK5	Winged-helix DNA-binding transcription factor family protein	229	ref|NP_175826.2 gb|ABE65711.1 gb|AEE33070.1	winged-helix DNA-binding transcription factor family protein; hypothetical protein At1g54240;	-	-	-	-	mature pollen grain, pollen tubes, cotylodones of heart stage embryo, weak induction by osmotic and heat stress, weak expression in guard cells and aba1 hypocotyl
25		*AT1G54240*		A0MEC6	Putative uncharacterized protein	230	gb|ABK28439.1	unknown					
26		*AT1G54240*		Q9SLK8	Putative uncharacterized protein F20D21.6	207	gb|AAD25606.1|AC005287_8	Hypothetical protein					
27	**GH1-Myb-related8** (SMH15)	AT1G54260.1	winged-helix DNA-binding transcription factor family protein	F4HV94	Winged-helix DNA-binding transcription factor family protein	197	ref|NP_175828.1| gb|AEE33072.1|	winged-helix DNA-binding transcription factor family protein	SMH15	-	-	-	mature pollen
28		*AT1G54260*		Q67YM4	Putative uncharacterized protein At1g54260	169	dbj|BAD44207.1|	hypothetical protein					
29		*AT1G54260*		Q9SLK7	Putative uncharacterized protein F20D21.8	227	gb|AAD25607.1|AC005287_9	Hypothetical protein					

Supplementary Table S2. List of articles referring to the role of plant linker histones.

Development			
General	Arabidopsis	Downregulation of all three Arabidopsis H1 variants (RNAi) leads to pleiotropic developmental defects at the vegetative and reproductive stages and impaired DNA methylation profiles	(Wierzbicki and Jerzmanowski 2005)
Meiosis	Tobacco	A 4-fold reduction of H1A and H1B levels impairs male meiosis and pollen development.	(Prymakowska-Bosak, Przewloka et al. 1999)
Endosperm development	Maize	H1/DNA ratio levels decrease during endoreduplication in maize endosperm in parallel to massive expression of storage genes	(Zhao and Grafi 2000)
Cell fate	Arabidopsis	H1.1 and H1.2 somatic variants are evicted in male and female meiotic precursors cells, transiently restored at meiosis and undetectable again in the functional megaspore.	(She, Grimanelli et al. 2013) (She and Baroux 2015)
Differentiation	Maize	H1 variants' ratios are dynamically regulated along the division and differentiation zones of maize root. Notably, the $H1^0$ variant increases in differentiation while H1A/H1B decrease	(Alatzas, Srebreva et al. 2008)
Seed biology	Maize	GWAS association with seed composition traits identified H1 loci with starch, protein and oil content	(Cook, McMullen et al. 2012)
	Maize	Onset of grain filling is associated with a change in properties of linker histone variants in maize kernels	(Kalamajka, Finnie et al. 2010)
	Rapeseed	Osmopriming (exogenous control of seed imbibition) and seed germination correlate with decreased levels of H1 mRNAs in *Brassica oleracea*	(Soeda, Konings et al. 2005)
Fruit ripening	Banana	Fruit ripening and ethylene treatment increases the MaHIS1 H1 variant (homologous to the Arabidopsis H1.1 variant) in *Musa acuminate*	(Wang, Kuang et al. 2012)
Biotic/abiotic stress			
Drought	Tomato	H1-S variant is up-regulated under water deficit conditions. Antisense-mediated downregulation suggests a role of H1-S in plant water status regulation and stomatal functions.	(Scippa, Griffiths et al. 2000) (Scippa, Di Michele et al. 2004)
Drought	Arabidopsis	The stress-inducible H1.3 variant is distinct from H1.1 and H1.2 by is short C-terminal tail, few amino acid substitution in the binding domain and very high mobility. H1.3 is induced by combined light and water deficit and functions in stress responses and stomatal functions.	(Ascenzi and Gantt 1997) (Ascenzi and Gantt 1999) (Rutowicz, Puzio et al. 2015)
Drought	Cotton	Identification by mass spectrometry of a stress-inducible H1 variant in a drought tolerant cultivar (Vagad). This variant is absent from the drought sensitive cultivar RAHS-14.	(Trivedi, Ranjan et al. 2012)
Various biotic and abiotic stresses	Banana	Chilling or exogenous application of methyljasmonate, H2O2 or ABA induced MaHIS1 (homologous to *At*H1.1) mRNA levels transiently. Exposure to the fungal pathogen *Colletotrichum musae* induced a prolonged increase in MaHIS1 mRNA levels.	(Wang, Kuang et al. 2012)
Epigenetic regulation			
DNA methylation	Arabidopsis	RNAi downregulation of the three H1 variants led to local fluctuations in DNA methylation patterns in both CG and non-CG contexts .	(Wierzbicki and Jerzmanowski 2005)
DNA methylation	Arabidopsis	Loss-of-function of the three main H1 variants causes hypermethylation at heterochromatic transposons and partially rescues the hypomethylation phenotype of *DECREASED IN DNA METHYLATION1* (*ddm1*) mutants	(Zemach, Kim et al. 2013)
Imprinting	Arabidopsis	H1 variants interacts with the DNA glycosylase DEMETER (yeast two hybrid and GST pulldown assays). H1 depletion reduces maternal expression of DME target genes (*MEA, FWA, FIS2*) in correlation with increased DNA methylation levels.	(Rea, Zheng et al. 2012)
Histone deacetylation	Arabidopsis	H1 directly interacts with the Histone Deacetylase Complex 1 HDC1.	(Perrella, Carr et al. 2016)
Transcriptional regulation			
Lignin biosynthesis	Eucalyptus	H1.3 interacts with the transcription factor MYB1 and contributes to transcriptional repression of genes involved in lignin biosynthesis.	(Soler, Plasencia et al. 2016)

Enhances TF binding	Rice/wheat	H1 facilitates binding of the transcription factor EmBP-1 to the ABA-responsive gene *Em*.	(Schultz, Spiker et al. 1996)
Regulates stress-responsive genes	Arabidopsis	H1.3 contributes to induce stress-response associated factors under combined light and drought stress	(Rutowicz, Puzio et al. 2015)
Structural function			
Chromatin condensation	Pea	Lower chromatin condensation in callus cells compared to root cells correlate with varying levels of histone H1 variants	(Bers, Singh et al. 1992)
	Tobacco	Overexpression of an Arabidopsis H1 variants in tobacco induces strong heterochromatinization	(ŚLUSARCZYK, PRYMAKOWSKA-BOSAK et al. 1999)
	Pea, Maize, Bean	The proportion of extracted H1 correlates with the level of genomic repeats and the degree of chromatin condensation (transmission electron microscopy)	Oleszweska, 1988
Other cellular functions			
Microtubule organization	Tobacco	In tobacco BY-2 cells, H1B functions as a microtubule-organizing factor on the nuclear surface showing DNA independent functions. Probably interacting with tubulin.	(Hotta, Haraguchi et al. 2007) (Nakayama, Ishii et al. 2008) (Kaczanowski and Jerzmanowski 2001)

References

Alatzas, A., L. Srebreva, et al. (2008). "Distribution of linker histone variants during plant cell differentiation in the developmental zones of the maize root, dedifferentiation in callus culture after auxin treatment." Biol Res **41**(2): 205-215.

Ascenzi, R. and J. S. Gantt (1997). "A drought-stress-inducible histone gene in Arabidopsis thaliana is a member of a distinct class of plant linker histone variants." Plant Mol Biol **34**(4): 629-641.

Ascenzi, R. and J. S. Gantt (1999). "Molecular genetic analysis of the drought-inducible linker histone variant in Arabidopsis thaliana." Plant Mol Biol **41**(2): 159-169.

Ascenzi, R. and J. S. Gantt (1999). "Subnuclear distribution of the entire complement of linker histone variants in Arabidopsis thaliana." Chromosoma **108**(6): 345-355.

Bers, E. P., N. P. Singh, et al. (1992). "Nucleosomal structure and histone H1 subfractional composition of pea (Pisum sativum) root nodules, radicles and callus chromatin." Plant Mol Biol **20**(6): 1089-1096.

Cook, J. P., M. D. McMullen, et al. (2012). "Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels." Plant Physiol **158**(2): 824-834.

Hotta, T., T. Haraguchi, et al. (2007). "A novel function of plant histone H1: microtubule nucleation and continuous plus end association." Cell Struct Funct **32**(2): 79-87.

Kaczanowski, S. and A. Jerzmanowski (2001). "Evolutionary correlation between linker histones and microtubular structures." J Mol Evol **53**(1): 19-30.

Kalamajka, R., C. Finnie, et al. (2010). "Onset of grain filling is associated with a change in properties of linker histone variants in maize kernels." Planta **231**(5): 1127-1135.

Nakayama, T., T. Ishii, et al. (2008). "Radial microtubule organization by histone H1 on nuclei of cultured tobacco BY-2 cells." J Biol Chem **283**(24): 16632-16640.

Perrella, G., C. Carr, et al. (2016). "The Histone Deacetylase Complex 1 Protein of Arabidopsis Has the Capacity to Interact with Multiple Proteins Including Histone 3-Binding Proteins and Histone 1 Variants." Plant Physiol **171**(1): 62-70.

Prymakowska-Bosak, M., M. R. Przewloka, et al. (1999). "Linker histones play a role in male meiosis and the development of pollen grains in tobacco." Plant Cell **11**(12): 2317-2329.

Rea, M., W. Zheng, et al. (2012). "Histone H1 affects gene imprinting and DNA methylation in Arabidopsis." Plant J **71**(5): 776-786.

Rutowicz, K., M. Puzio, et al. (2015). "A Specialized Histone H1 Variant Is Required for Adaptive Responses to Complex Abiotic Stress and Related DNA Methylation in Arabidopsis." Plant Physiol **169**(3): 2080-2101.

Schultz, T. F., S. Spiker, et al. (1996). "Histone H1 enhances the DNA binding activity of the transcription factor EmBP-1." J Biol Chem **271**(42): 25742-25745.

Scippa, G. S., M. Di Michele, et al. (2004). "The histone-like protein H1-S and the response of tomato leaves to water deficit." J Exp Bot **55**(394): 99-109.

Scippa, G. S., A. Griffiths, et al. (2000). "The H1 histone variant of tomato, H1-S, is targeted to the nucleus and accumulates in chromatin in response to water-deficit stress." Planta **211**(2): 173-181.

She, W. and C. Baroux (2015). "Chromatin dynamics in pollen mother cells underpin a common scenario at the somatic-to-reproductive fate transition of both the male and female lineages in Arabidopsis." Front Plant Sci **6**: 294.

She, W., D. Grimanelli, et al. (2013). "Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants." Development **140**(19): 4008-4019.

ŚLUSARCZYK, J., M. PRYMAKOWSKA-BOSAK, et al. (1999). "Ultrastructural Organization of Leaves of Transgenic Tobacco Overexpressing Histone H1 from Arabidopsis thaliana." Annals of Botany **84**(3): 329-335.

Soeda, Y., M. C. Konings, et al. (2005). "Gene expression programs during Brassica oleracea seed maturation, osmopriming, and germination are indicators of progression of the germination process and the stress tolerance level." Plant Physiol **137**(1): 354-368.

Soler, M., A. Plasencia, et al. (2016). "The Eucalyptus linker histone variant EgH1.3 cooperates with the transcription factor EgMYB1 to control lignin biosynthesis during wood formation." New Phytol.

Trivedi, I., A. Ranjan, et al. (2012). "The histone H1 variant accumulates in response to water stress in the drought tolerant genotype of Gossypium herbaceum L." Protein J **31**(6): 477-486.

Wang, J. N., J. F. Kuang, et al. (2012). "Expression profiles of a banana fruit linker histone H1 gene MaHIS1 and its interaction with a WRKY transcription factor." Plant Cell Rep **31**(8): 1485-1494.

Wierzbicki, A. T. and A. Jerzmanowski (2005). "Suppression of histone H1 genes in Arabidopsis results in heritable developmental defects and stochastic changes in DNA methylation." Genetics **169**(2): 997-1008.

Zemach, A., M. Y. Kim, et al. (2013). "The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin." Cell **153**(1): 193-205.

Zhao, J. and G. Grafi (2000). "The high mobility group I/Y protein is hypophosphorylated in endoreduplicating maize endosperm cells and is involved in alleviating histone H1-mediated transcriptional repression." J Biol Chem **275**(35): 27494-27499.