Zn$^{2+}$-dependent histone deacetylases are widely distributed in archaea, bacteria, and eukaryotes. Through deacetylation of histones and other biomolecules, these enzymes regulate mammalian gene expression, microtubule stability, and polyamine metabolism. In plants, they play essential roles in development and stress response, but little is known about their biochemistry. We provide here a holistic revision of plant histone deacetylase (HDA) phylogeny and translate recent lessons from other organisms. HDA evolution correlates with a gain of structural ductility/disorder, as observed for other proteins. We also highlight two recently identified Brassicaceae-specific HDAs, as well as unprecedented key mutations that would affect the catalytic activity of individual HDAs. This revised phylogeny will contextualize future studies and illuminate research on plant development and adaptation.

Diversification of Histone Deacetylases

Histone deacetylases (see Glossary, named HDACs in mammals and HDAs in plants) are widely distributed in all kingdoms of life. These enzymes regulate gene expression among other biological processes, play a crucial role in development and cell-cycle progression, and their deregulation is associated with human diseases such as cancer and neurodegeneration [1–3].

In eukaryotes, nuclear HDACs deacetylate the histone proteins that scaffold DNA packing into nucleosomes, rendering them key epigenetic regulators [4–6]. Histone deacetylation leads to stronger histone–DNA interactions, stimulates chromatin condensation, and thus triggers transcriptional repression [7]. By contrast, HDACs delete the N$^{ε}$-acetyl-lysine recognition motif for bromodomain-containing transcriptional coeffectors and activators [8]. Overall, given the opposing effects of histone acetyltransferases (HATs), there is a strict balance between acetylation and deacetylation of histones that constitutes a pivotal regulatory mechanism of gene expression [8]. Importantly, HDACs regulate the function of biomolecules other than histones, both in the nucleus and in the cytoplasm. For example, HDAC3 deacetylates phosphoglycerate kinase (PGK) [9], HDAC6 regulates microtubule dynamics via deacetylation of α-tubulin [10,11], and HDAC10 takes part in the metabolism of polyamines by deacetylating N$^{ε}$-acetylsperrmidine [12].

In plants, reversible histone acetylation affects genome stability, transcriptional regulation, development, and stress responses [13,14]. In addition, HDAs regulate the cell cycle and programmed cell death, in part via interaction with key transcription factors [15,16], and they have also been shown to deacetylate polyamines other than histones [17].

There are three families of lysine deacetylases in eukaryotes, based on three conserved deacetylase domains that rely on different cofactors [18]. HDAs refer here to the Rpd3/Hda1 (reduced potassium-dependence 3/histone deacetylase 1) family of Zn$^{2+}$-dependent hydrolases, which share similarity to yeast Rpd3, further grouped into classes I, II, and IV [19,20]. Phylogenetic analyses suggest that these three HDAC/HDA classes are constituted by ancient enzymes that evolve from bacterial
HDAC-related proteins [19]. The sir2 (silent information regulator 2) family of NAD⁺-dependent enzymes (the sirtuins, class III HDACs) and the plant-specific HD2 family (or HD-tuins) possess catalytic domains unrelated to those of the Rpd3/Hda1 family [21,22] and are not the focus of this review.

The diversification of Zn²⁺-dependent deacetylases in all kingdoms of life indicates a high degree of evolutionary plasticity, and suggests that each of the HDA classes has a non-redundant and essential role. Plant Rpd3 homologs were first identified in maize in 1998 [23], and have been since characterized in arabidopsis (Arabidopsis thaliana), Zea mays (maize), Oryza sativa (rice), and other model and crop plant species [24–30]. Various HDA phylogenetic analyses have been published, including a limited number of species or focused on individual genomes (Box 1) [25,27,29–31]. In this review we provide a holistic evolutionary revision of the plant HDA family, including 60 species representative of 42 families (Figure S1 in the supplemental.

Box 1. History of the Phylogeny of Plant HDAs

Plant Zn²⁺-dependent HDAs are divided into classes I, II, and IV, and they can be further grouped according to similarity to aradiopsis sequences [25,32]. Pandey et al. identified two clusters within class I containing the AthDA19 (also known as HDA1), and AthDA6 and AthDA7 genes, respectively (Figure IA). Each cluster included dicot and monocot genes, suggesting a genome duplication origin before the divergence of the two lineages [32]. In addition, in the phylogeny of Pandey et al. class II HDAs include the AthDA5, AthDA15, and AthDA18 genes, and class IV contains AthDA2, which has similarity to bacterial Aac and cyanobacterial glutamine synthetase [32]. Note that class IV was earlier designated as class III, but was then renamed class IV to distinguish it from the sir2 family. In this study, arabidopsis sequences [25,32], Pandey et al. proposed a revised class II that included AthDA5, AthDA8, AthDA14, AthDA15, and AthDA18 (Figure IB) [25]. In other studies, such as those conducted by Fu et al. in Oryza sativa, it has been suggested that the Rpd3/Hda1 HDA family could be divided into four alternative classes [117]. This review allows the elaboration of a revised HDA phylogenetic tree (Figure IC) and the identification of key HDA variants with potential alternative functions. We conclude that AthDA7 and AthDA18 do not constitute canonical subclasses of HDAs, and are instead members of subclasses HD6 of class I and HD5 of class II HDAs, respectively (hereafter HD6-2 and HD5-2). They are associated with Cyanobacteria lineage-specific genes that derived from a recent divergence event. In addition, the HDA9 subclass of class I is split into two subgroups, HDA9a and HDA9b, as a result of evolutionary divergence in the monocot lineage.

(A) Class I
HDA7
HDA8
HDA10
HDA11
HDA15
HDA14
HDA19
HDA21
HDA22

(B) Class II
HDA1
HDA2
HDA4
HDA5
HDA6
HDA9
HDA10
HDA12
HDA13
HDA14
HDA15
HDA18

(C) Class III
HDA1
HDA2
HDA4
HDA5
HDA6
HDA8
HDA10
HDA12
HDA13
HDA14
HDA15
HDA16
HDA17
HDA19
HDA21
HDA22

Figure 1. Diversification and Phylogeny Revision of Plant HDA Classes. Schemes of the phylogenetic cladograms proposed for plant Zn²⁺-dependent HDAs by (A) [32], (B) [26], and (C) this review. *Truncated sequences that lack the HDA catalytic pocket. **Cyanobacteria-specific HDAs; HD6-2 was previously named HDA7 and HD6-2 was termed HDA18. ***Monocot-specific subclass. Cladograms of panels (A) and (B) are adapted, with permission, from [32] and [25].

Glossary

Angiosperms: flowering plants, vascular plants that are distinguished by the production of seeds within an enclosed ovary. These constitute the most abundant and diverse group of land plants.

Arecaceae, Bromeliaceae, Musaceae, Zostereaceae: families of monocotyledonous plants.

Brassicaceae: a family of dicots that includes economically important herbaceous plants. Their diversification is dated to ~100 Mya.

Bryophytes: non-vascular land plants that include liverworts, hornworts, and mosses. They belong to an old lineage dated to ~500 Mya.

Chlorophytes: ‘green algae’, a highly paraphyletic group of green plants that mostly live in aquatic environments. Their origin is dated to ~1200 Mya.

Cladogram: ‘evolutionary tree’, a diagram that shows predicted phylogenetic relationships among organisms, genes, or proteins. Cladograms use lines to distinguish between paraphylogeny and orthology during the course of evolution.

Cyanobacteria: blue–green bacteria, a phylum of photosynthetic prokaryotes. They were involved in the origin of primary plastids and chloroplasts through endosymbiotic events, and their origin is dated to 2700–1400 Mya.

Dicots: dicotyledons, a paraphyletic group of angiosperms that are distinguished by seeds that have two embryonic leaves or cotyledons. The divergence between monocots and dicots occurred at 145–230 Mya.

Divergence: also known as divergent evolution, the accumulation of differences between closely related populations within a species leading to speciation. This term can also be applied in molecular evolution when proteins derive from homologous genes, either orthologs or paralogs.

Ductile region: an intrinsically disordered region (IDR) of a protein that does not fold into a well-defined 3D structure. Ductile regions can be either entirely disordered or partially disordered, spanning only a few contiguous disordered residues (<10 aa) or containing long segments (≥30 aa) of continuously disordered residues. Algorithms have been developed that can predict the likelihood that an amino acid in a protein sequence is structured or disordered [42–44].
information online). This allows the identification of key HDA variants with potential alternative functions. Furthermore, we discuss structural differences in the context of the recent literature on mammalian HDACs.

**Phylogeny and Evolution of Plant HDAs**

The extensive genomic information currently available on families of plants from chlorophytes to angiosperms enables precise revision of plant HDA phylogeny. Importantly, we verified the sequence composition of each protein so as to discard incomplete proteins and identify potentially inactive enzymes that lack essential amino acid residues.

**Class I HDAs**

Two major distinct clades are distinguished in dicots (Figure 1A) and monocots (Figure 1B): clade 1 with subclass HDA9, and clade 2 with subclasses HDA19 (also named HDA1, hereafter HDA1/19) and HDA6. Subclasses HDA1/19 and HDA6 share a common ancestor and have evolved independently of subclass HDA9. As expected, orthologs in Brassicaceae are clustered (bootstrap >85) in the three subclasses according to their phylogeny (Figure 1A).

Earlier studies in arabidopsis reported a distinct ortholog, named HDA7 (At5g35600), which could belong to a fourth subclass of class I [25,32]. AthaHDA7 and subsequent orthologs from the Brassicaceae species Crub (pink shepherd’s purse), Bnap (colza), Bole (broccoli, cabbage), and Brap (napa cabbage) are found within subclass HDA6 and are positioned close to AthaHDA6 (At5g63110), which indicates that they are paralogs (Figure 1A, light-blue circles). Furthermore, HDA7-type orthologs are not present in other related phylogenetic families, indicating that HDA7 is encoded by Brassicaceae lineage-specific genes that are associated with a recent divergence event. In fact, Brassica genomes have undergone whole-genome duplication (WGD) and lineage-specific whole-genome triplication (WGT) events followed by diploidization [33–35]. For clarity in the annotation, we recommend that the paralogs of AthaHDA6 and AthaHDA7 (and their respective orthologs) are renamed HDA6-1 and HDA6-2, respectively (Figure 1A).

Chlorophyte homologs are only present in clade 1 within subclass HDA9 (bootstrap >98, Figure 1), indicating that subclass HDA9 has a more ancient evolutionary origin than subclasses HDA1/19 and HDA6, and that these originated before the divergence of the monocot and dicot lineages. In addition, two distant paralysic subgroups (bootstrap >90) appear in the monocot subclass HDA9 (Figure 1B, named HDA9a and HDA9b). Interestingly, homologs from chlorophytes and the ancient freshwater alga Chara braunii (Cbra, ca. 450 Mya) [36,37] cluster close to the HDA9a subgroup (bootstrap >95) in dicot and monocot trees. This suggests that subgroup HDA9b diverged later as a result of a WGD event in the monocot lineage. Note that HDA9b orthologs in the monocot lineage are often annotated as HDA6 (i.e., Acorn, XP_020098843.1; Atau, XP_020188071.1; Doli, OEL25058; Pdac, XP_008791160.1; Sbic, XP_002444249.1), which can lead to confusion because HDA6 constitutes a different subclass of class I. In the dicot phylogeny (Figure 1A), HDA9 of Csub clustered together with green algae homologs, which could indicate an ancestral origin for this protein, but it could also reflect errors in the reported sequence.

**Class II HDAs**

Class II HDAs comprise two major clades: clade 1, with subclasses HDA8 and HDA14; and clade 2, with subclasses HDA5 and HDA15 (Figure 2). Subclasses HDA8 and HDA14 are only present in plants but still exhibit high sequence similarity to the conserved catalytic domains of HDA5 and HDA15, which are common to all eukaryotic organisms [26]. Thus, subclasses HDA8 and HDA14 likely diverged separately from HDA5 and HDA15.
(A) Subclass HDA6

HDA6-1

HDA6-2

(B) Subclass HDA9

HDA9a

HDA9b

Chlorophyte

Bryophyte

Dicot

Monocot

Trends in Plant Science

(See figure legend at the bottom of the next page.)
By analogy to AthaHDA7, previous reports revealed a distinct ortholog within class II named HDA18 (At5g61070) that was assigned to a different subclass [25,32]. The AthaHDA18 protein (Figure 2A, light-blue circles) is found here in the same monophyletic group (bootstrap >95) as AthaHDA5 (At5g61060), indicating that AthaHDA18 and AthaHDA5 are paralogs. In addition, orthologs of AthaHDA18 are identified in other Brassicaceae species such as Crub (pink shepherd’s purse), Bnap (colza), Bole (cabbage) and Brap (napa cabbage), which suggests that HDA18 is also encoded by a Brassicaceae lineage-specific gene that is associated with a recent divergence event. For clarity in the annotation, we recommend that AthaHDA5 and AthaHDA18 and their respective orthologs are renamed HDA5-1 and HDA5-2, respectively (Figure 2A). The existence of these paralogs of HDA6 (class I) and HDA5 (class II) could help in understanding the divergence and adaptation mechanisms in Brassica plants.

Class II Brassicaceae and Poaceae HDAs form separated monophyletic subgroups (bootstrap >98) in the HDA5, HDA8 clades, and in the HDA14 clade of Brassicaceae and the HDA15 clade of Poaceae, following the angiosperm phylogeny. An exception to this is OsatHDA5, which is grouped with Arecacea, Bromeliaceae, and Musacea members (bootstrap >85). Furthermore, the dicot tree suggests that subclasses HDA8, HDA14, and HDA15 had an earlier ancestor in this lineage, and that they share a distant evolutionary relationship compared to subclass HDAs (Figure 2A).

Lastly, some HDA5 orthologs from the distant Papaveraceae family (with respect to Brassicaceae) are annotated as HDA18-like (i.e., Ppsom, XP_026381159), and some HDA14 orthologs in monocot lineage are annotated as HDA10 (i.e., Zmay, XP_008673398), which can lead to confusion. The terms HDA10 and HDA17 have been previously associated with truncated members of class I HDAs in arabidopsis [25,32]. However, the truncation has eliminated a large portion of the HDA catalytic domain, including key catalytic residues, indicating that these proteins have a different structure and function [25]. Therefore, we consider it more accurate to not include HDA10 and HDA17 in the plant HDA family tree (Box 1).

Class IV HDAs
Class IV HDAs are distributed into two major distinct clades: clade 1 with chlorophyte, bryophyte, lycophyte, and monocot sequences, and clade 2 with dicot sequences (Figure 3A). This clustering pattern is consistent with the diversification of the two lineages during evolution. In clade 1,
monocot Poaceae sequences are separated into two clusters (bootstrap >90), and Arecaceae, Bromeliaceae, Musaceae, and Zosteraceae form distinct clusters close to bryophytes and chlorophytes, suggesting a common ancestor. Clade 2 follows the clustering pattern of dicot phylogeny, including separated clusters for the ancestral dicots Acoe, Nnuc, and Psom (bootstrap >90).

Evolution and Ductile Regions in the HDA Family
It is well documented that protein ductility increases in concert with organismal complexity and multicellularity [38,39]. This means that the proportion of flexible or ductile regions (often named intrinsically disordered regions, IDRs) in protein sequences generally increases in eukaryotes relative to bacterial organisms. It is predicted that prokaryotic proteomes contain 12–30% residues within long IDRs, whereas in eukaryotes this prediction is 33–50% [40–42]. This gain of IDRs in eukaryotes is hypothesized to be associated with differences in the genetic machinery [43]. Importantly, individual eukaryotic proteins with an endosymbiotic origin may maintain bacterial characteristics in their sequences [43].

Class I and II HDACs include characteristic long IDRs at the N-terminus and/or C-terminus [20], which are often indicative of protein interaction networks [44]. In class I, longer IDRs are predicted in the C-termini of plant subclasses HDA1/19 (150–160 amino acids, aa), HDA6 (80–100 aa), and the monocot subclass HDA9b (80–100 aa) compared to subclass HDA9a (40–50 aa). Accordingly, HDA9a sequences contain ~17% disordered residues, a percentage much lower than those predicted in HDA1/19, HDA6, and HDA9b (30–35%, Figure 3B), consistent with their ancient origin. Moreover, HDA9a sequences present a high similarity (up to 60%) with bacterial proteins (i.e., GeneBank ID: HHG08807.1, from metagenome analysis in Yellowstone National Park). In Brassicaceae, HDA6-2 orthologs lack the long flexible C-terminal tail that is characteristic of subclass HDA6, explaining the reduction in disordered residue fraction from ~33% to ~22%. In class II, subclass HDA15 presents a long IDR in the N-terminus (ca. 130 aa) that is not predicted in the other subclasses, thus increasing disordered residues to ~33% (Figure 3B) and suggesting an earlier eukaryotic origin. Similarly, some Brassicaceae HDA5-2 orthologs present up to 50% disordered residues, suggesting that they have a more recent origin than HDA5-1 and a clear increase in plasticity. Note, Brassica species underwent recent WGD and hexaploidization events (23–47 Mya), and this affected the evolution of many agriculturally important species [45,46].

Most cyanobacterial nuclear genomes contain two histone deacetylases that are homologs of classes II and IV, whereas no HDAs are encoded in the chloroplast genome. Cyanobacterial HDAs present high similarity (47–66%) to plant HDA14 and HDA2, respectively (Figure S2)
Figure 3. Phylogeny of Plant Class IV Histone Deacetylases (HDAs) and Ductility/Disorder in HDA Evolution.

(A) Phylogenetic cladogram of plant class IV HDAs including 31 families, 51 species, and 51 sequences. Sequences belong to chlorophyte (Ceus, Mpus, Vcar), bryophyte (Ppat), lycophyte (Smoe), and monocot and dicot families: Actinidiaceae (Acth), Amarantaceae (Bvul), Arecaceae (Pdac), Asteraceae (Hann), Bromeliaceae (Acim), Brassicaceae (Atha, Bnap, Bole, Brap, Crub), Cannabaceae (Pand), Characeae (ancestral plant Cbra), Cucurbitaceae (Csat), Euphorbiaceae (Manhot scutelora, Mesc; R. communis, Rcom), Fabaceae (Lotus japonicus, Ljap, Pvl, Palt), Fabaceae (Oeub), Lythraceae (Pgra), Malvaceae (Grai), Moraceae (Morus notabilis, Mnor), Musaceae (Macu), Myrtaceae (Eucalyptus grandis, Egna), Nelumbonaceae (ancestral eudicot Nnu), Papaveraceae (ancestral eudicot Psorn), Poaceae (Atau, Bdis, Doli, Hvil, Osat, Phal, Sbic, Sita, Zmay), Ranunculaceae (ancestral eudicot Acce), Rubicaceae (Ccarn), Rutaceae (Ccile), Salicaceae (Pntr), Solanaceae (Slyc), Theaceae (Csit), Vitaceae (Vvin), and Zosteraceae (Zostera marina, Zmar). Outgroup: AthaHDA11/19. Analyses were performed as described in Figure 1. (B) Box-plot distribution of the percentage of disordered residues in HDA proteins from cyanobacteria and plants, based on predictions from PONDR VSL2b algorithm [119,120]. (C) Phylogenetic tree of HDAs of subclasses HDA2, HDA14, and HDA5 from cyanobacteria and plants. Sequences of cyanobacteria belong to: Cnncalum epipsmum PCC 9333 (Cep14, Cep2); Chromococciopsis thermalis PCC 7203 (Cthe14, Cthe2); Fischerella sp. NIES-4106 (Fisc14, Fisc2); Gloeothecia citriformis PCC 7424 (Gcit14.1, Gcit14.2, Gcit2); Gloecobacter violaceus PCC 7421 (Gvio14, Gvio2); Halomicronema hongdechloris C2206 (Hhon14, Hhon2); Leptolyngbya boryana dg5 (Lbor14, Lbor2); Microcystis aeruginosa NIES-843 (Maer14); Nostoc punctiforme PCC 73102 (Npunc14, Npunc2); Oscillatoria acuminata PCC 6304 (Oacu14, Oacu2); Stauraria cyanoxiphina PCC 7437 (Socs14); Synechococcus elongatus PCC 11801 (Sync14, Sync2); and Synechocystis sp. PCC 6803 (Syn14).
which contain 14–20% of total disordered residues and are therefore closer to the prokaryotic proteome average (Figure 3B). This indicates that the HDA2 and HDA14 plant proteins conserve a sequence fingerprint related to their bacterial origin. Previous reports suggested that the relationship between eukaryote HDAC/HDAs and bacterial homologs is uncertain, and that multiple horizontal transfer events might have taken place [19]. Interestingly, the cladogram in Figure 3C shows that plant HD2 and HDA14 are closely related to cyanobacterial histone deacetylase proteins (bootstrap >78), which could suggest that plant HDA genes were incorporated via endosymbiosis and subsequent transfer from the bacterial plasmid to the nuclear genome (plastid origin). Not surprisingly, 800–2000 genes in the arabidopsis genome are believed to have a cyanobacterial origin [47,48]. Moreover, HDA14 has been localized in the chloroplast in arabidopsis, where it deacetylates photosynthetic proteins [17].

The Zn\(^{2+}\)-Dependent Active Site in Plant HDAs

The sequence organization of HDACs/HDAs has been extensively reviewed [20,49–51]. In plants, bioinformatics tools have enabled prediction of HDA sequence architectures, as displayed in the Pfam and Interpro databases [25,29]. Class I, II, and IV HDAs harbor the conserved histone deacetylase domain (313–330 aa), which covers 60–75% of the entire sequence, and often a nuclear localization signal (NLS) at the N-terminus. An additional C-terminal NLS motif can be observed in class I HDA1/19 and class II HDA15. Conversely, the class II enzymes HDA8, HDA15, the single Brassicaceae HDA5-2 (before HDA18), and the class IV member HDA2 all contain nuclear export signal (NES) motifs [25]. HDA15 contains a zinc-finger RanBP2 domain located close to the N-terminus, which is also found in mammalian HDAC6 and has been linked to ubiquitin binding and targeting for lysosomal degradation [52,53].

The HDAC/HDA catalytic domain has been characterized by analysis of several crystal structures [51,54,55] which helped to identify residues that are essential for Zn\(^{2+}\) cofactor binding and amide-bond hydrolysis (Figure 4A) [54–57]. In brief, the Zn\(^{2+}\) ion is bound to contiguous Asp and His residues, and to an additional Asp residue in a separate loop. Once the N\(^{\varepsilon}\)-acetyl-lysine substrate is bound in the active site, a water molecule attacks the amide and forms a tetrahedral intermediate that is stabilized by a Tyr residue (Figure 4A). This step is also catalyzed by two His-Asp dyads, of which one appears to serve as a proton-exchange catalyst whereas the other helps to stabilize the transition state [57]. The cycle concludes with formation of the Lys and acetate products, assisted by the protonated His-Asp dyad (Figure 4A). Most HDA protein sequences analyzed in this review conserve these essential residues. However, some exceptions in each class are worth discussing.

Class I HDAs

Class I plant HDAs are, on average, ~470 residues in length, and they share up to 58–68% sequence identity with human class I HDACs 1–3 and 40–50% with HDAC8. Plant sequences conserve (i) the catalytic residues His148, His149, and Tyr311 (referred to AtaHDA1/19, At4g38130, Uniprot O22446), that are equivalents to His140, His141, and Tyr303 in human HDAC1 (Uniprot Q13547) [56–58], in which the catalytic Tyr is preceded by a conserved poly (Gly) motif; and (ii) the Zn\(^{2+}\)-binding residues Asp184, His186, and Asp272 that are equivalents to Asp176, His178, and Asp264 [58]. Surprisingly, the HDA1/19 orthologs Atau M8BTX9, Macu M0SW61, and Macu M0S9F0 are exceptions to these rules because they have lost the catalytic residues equivalent to His148 and His149, indicating complete loss of deacetylase activity (Figure 1B, black circles).

Of particular note are mutations that are present in Brassicaceae HDA6-2 orthologs (before HDA7, Figure 1A, light-blue circles), such as substitution of the catalytic Tyr residue by His in
Figure 4. 3D Structural Alignment and Key Active-Site Mutations In Histone Deacetylases. (A) Mechanism of catalysis of Zn²⁺-dependent histone deacetylases, showing residues involved in Zn²⁺ binding, transition state stabilization, and proton transfer [57]. The last step involves two proton transfers, one proton from the His-Asp dyad and a second proton from the acetate moiety. A single His-Asp dyad is shown for simplicity. (B) Superposition of AthaHDA1/19 (O22446, e-value 4.2 × 10⁻⁸⁸) with the X-ray crystal structure of Homo sapiens (Hsap) HDAC1 (PDB 4BKX, [58]). Amino acid positions are, for HsapHDAC1: His140 (#1), His141 (#2), Asp176 (#1), His178 (#3), Asp264 (#2), and Tyr303; and for AthaHDA1/19: His148 (#1), His149 (#2), Asp184 (#1), His186 (#2), Asp272 (#2), and Tyr311. (C) Superposition of BnapHDA6-2.2 (XP_013737742, e-value 7.3 × 10⁻⁸⁰) with the X-ray crystal structure of HsapHDAC1 (PDB 4BKX [58]). BnapHDA6-2.2 amino acid positions are: His150 (His/His), Arg151, Gly186, Tyr188 (His/Tyr), Asp272, and His311 (Tyr/His). Abbreviation: Cat., catalytic residue. (D) Section view of the active-site channel in AthaHDA1/19 and BnapHDA6-2.2. The conserved position of the Zn²⁺ ion is shown for BnapHDA6-2.2, although the lack of coordinating residues makes its presence unlikely. (E) Superposition of AtauHDA5.1 (XP_020157471, e-value 1.7 × 10⁻⁵²) with the X-ray crystal structure (Figure legend continued at the bottom of the next page.)
Bnap (HDA6-2.1: A0A078IK3 and HDA6-2.2: XP_0137377742), Bole (XP_013603394), and Brap (XP_033129960) (Figure S3). This modification resembles the active-site mutation that is characteristic of class IIa mammalian HDACs, and which abolishes deacetylase activity [59,60]. However, in these four HDA6-2 proteins it is accompanied by further mutations in the equivalents to the catalytic His149 (H-to-R), the Zn²⁺-binding residues Asp184 (D-to-G) and His186 (H-to-Y, only in Bole XP_013603394 and in Bnap XP_0137377742), and a Gly residue in the poly(Gly) motif (G-to-Q, Figure S3). Although there is no experimental data on the catalytic activity of HDA6-2 proteins, it appears unlikely that these four proteins act as deacylases. The HDA6-2 orthologs of Atha (Q9FH09) and Crub (XP_006293194) also contain mutations in the Zn²⁺-binding residues equivalent to Asp184 (D-to-G) and His186 (H-to-P). Therefore, we expect that the HDA6-2 proteins are not catalytically active histone deacetylase enzymes and that they might have adopted a different role in Brassicaceae biology. Multiple histone deacetylases exert their function in multiprotein complexes that assemble through protein–protein interactions [61,62]. In this regard, it is also remarkable that Brassicaceae HDA6-2 proteins lack the long flexible C-terminal tail that is characteristic of subclass HDA6, and this would likely reduce their ability to interact with protein partners. These features could have resulted from high selection pressure during plant speciation, affording a novel function. Future studies will assess whether HDA6-2 proteins fulfill different biological roles than HDA6-1 and earlier HDA6 members.

X-ray crystal structures of class I HDACs have been recently determined at high resolution from diverse bacteria and mammals, most of them fixed by cocrystallization through inhibitor binding [51,61]. These structural studies have been useful in understanding the configuration of the HDAC active site and the mechanism of catalysis [54]. No class I crystal structures have so far been reported in plants, but these structures could help to advance the plant HDA field through 3D structural alignment, similarly to how the bacterial HDLP structure launched research on mammalian HDACs [63].

Superposition of class I AthaHDA1/19 (O22446, Figure 4B) and of the non-canonical BnapHDA6-2.2 (XP_0137377742, Figure 4C) with human HDAC1 (PDB 4BKX, 67% and 49% identity, respectively) [58] reveals high correlation in the structure of the catalytic domain, which indicates that homology models could be built via thorough refinement. In fact, the predicted catalytic and Zn²⁺-binding residues in AthaHDA1/19 are at the exact positions of those in human HDAC1 (Figure 4B) [56,58,61]. The Zn²⁺ cofactor appears to be buried at the bottom of the characteristic 11 Å hydrophobic channel that accommodates the Nε-acetyl-lysine substrate [56,57]. This channel is extended into a ‘foot pocket’ that is only present in class I HDACs in mammals [51,64], but this could be because it was present in the crystal structure that was employed for alignment. In the case of BnapHDA6-2.2, the mutations in catalytic and Zn²⁺-binding residues completely reshape the active-site channel and would not accommodate the Zn²⁺ cofactor (Figure 4D). Thus, the 3D alignment supports a different biochemical activity and biological role for the Brassicaceae-specific HDA6-2 proteins.

Class II HDAs

Class II HDAs are, on average, ~500 residues in length and share up to 30–45% sequence identity with their human homologs (class IIA: HDACs 4, 5, 7, 9; class IIB: HDACs 6 and 10). All plant sequences have one catalytic domain, in contrast to mammalian class IIB HDACs, of AthaHDA15 (PDB 6J6T). Amino acid positions are, for AthaHDA15: His276 (#1), His277 (#2), Asp313 (#1), His315 (#3), Asp404 (#2), and Tyr444; and for AtauHDA5.1: His101 (#1), His102 (#2), Asp138 (#1), His140 (#2), Asp227 (#2), and Cys267. (F) Superposition of BnapHDA5-2 (A0A078HYV8, e-value 1.9 x e⁻47) with the X-ray crystal structure of AthaHDA15 (PDB 6J6T). BnapHDA5-2 amino acid positions are: His140 (#1), His141 (#2), Asp163 (#1), Asp228 (#2), Tyr268, and the ‘gatekeeper’ Arg235 (Leu411 in AthaHDA15). Superpositions were obtained from HHpred (https://toolkit.tuebingen.mpg.de/) [121,122].

Trends in Plant Science, Month 2020, Vol. xx, No. xx11
which HDAC6 has two catalytic domains and HDAC10 has one complete and one partial domain [51]. In plants, class II HDAs conserve (i) the catalytic residues His276, His277, and Tyr444 (referred to AthaHDA15, At3g18520, Uniprot F4J8S1) that are equivalent to His610, His611, and Tyr782 in the catalytic domain 2 (CD2) of human HDAC6 (Uniprot Q9UBN7) [65], and (ii) the Zn²⁺-binding residues Asp313, His315, and Asp404 that are equivalent to Asp649, His651, and Asp742. Surprisingly, the HDA5 ortholog of grass Atau (XP_020157471) contains a Cys residue in place of the catalytic Tyr, a difference that has not been reported to date in any organism (Figure 2B, red circle).

The catalytic site of most class II HDAs, similarly to class I and class IV, includes a Leu residue at the rim of the active site. This position has been found to be substituted by Lys in catalytic domain 1 (CD1) of mammalian HDAC6 [65], and by Gln in HDAC10 [12]. In these two cases the alternative amino acid serves as a ‘gatekeeper’ that provides substrate selectivity towards acetylated C-terminal Lys residues and polyamines, respectively [12,65]. In plants, Leu411 (referred to AthaHDA15) is changed to Gln in subclass HDA8, and to Arg in variants of Brassicaceae subclass HDA5 of Bnap (AOA078HYV8), Bole (VDD27184), and Brap (XP_009130142) (Figure 2A, red circles), but no Lys or Gln variants are found.

The sole 3D structure reported of a plant HDA is of the class II AthaHDA15 (PDB 6J6T, 2.36 Å resolution) [55]. This X-ray crystal structure shows tetrameric oligomerization for the histone deacetylase domain, and each monomer contains a typical ~10 Å channel with the catalytic Zn²⁺ ion at the bottom (Figure 4E). In addition, the N-terminal zinc-finger domain (aa 1–146) assists and stabilizes the dimerization of the histone deacetylase domain in vitro and in vivo, similarly to mammalian HDACs 1 and 2 [61,66], and is essential for its enzymatic activity [55]. As expected, the Zn²⁺ cofactor is bound to Asp313, His315, and Asp404, by analogy to HDAC6_CD2 (PDB 5EFN) [65], in which the predicted catalytic Tyr side chain is pointed towards it. Accordingly, Asp313, His315, and also His276 and His277, have been shown to be essential for the enzymatic activity of AthaHDA15 [55,67]. Superposition of the HDA5 ortholog of the grass Atau (XP_020157471, 60% identity) with the reported AthaHDA15 structure reveals that the active site Cys residue (Cys267) is in a position and orientation equivalent to that of the missing catalytic Tyr (Figure 4E). Because the corresponding Tyr is involved in stabilization of the tetrahedral intermediate formed during acetamide substrate hydrolysis [57], it seems unlikely that AtauHDA5 would be an active deacetylase, unlike close orthologs from arabidopsis and other species [69]. Superposition of BnapHDA5-2 (AOA078HYV8, 59% identity) with the structure of AthaHDA15 shows that the conserved Leu at the entrance of the active site is mutated to Arg (Arg234), and is positioned where it could act as ‘gatekeeper’ (Figure 4F). This could indicate that Brassicaceae HDA5-2 variants have a substrate specificity more similar to that of mammalian HDAC6_CD1, which harbors a Lys residue at the same position [65].

Class IV HDAs
Plant class IV HDAs are homologs of mammalian HDAC11 [19,25], are on average ~360 residues in length, and share up to 60% sequence identity with the human homolog. These sequences conserve (i) the catalytic residues His200, His201, and Tyr361 (referred to AthaHDA11, At5g26040) that are equivalent to His141, His142, and Tyr304 in human HDAC11 (Uniprot Q96DB2), and (ii) the Zn²⁺-binding residues Asp238, His240, and Asp318, the equivalents of Asp181, His183, and Asp261. Remarkable exceptions are the monocot Doli (OEL35364) and the non-legume symbiotic Pand (PON76711) orthologs that harbor mutations of the catalytic Tyr residues (Tyr267Asp and Tyr323Arg, respectively). These mutations in the active site have not been reported previously, and would likely cause loss of deacetylase activity for these enzymes or indicate an alternative catalytic mechanism. Interestingly, Pand diverged from legumes >100 million years ago and is capable of fixing N₂ via symbiosis [69].
### Biological Function of HDAs in Plants

Plant HDAs were first extracted in 1988 [70], and their inhibition was key for studying plant histone acetylation in the 1990s [71–73]. Studies since then have focused on their mRNA expression, subcellular localization, and the elucidation of their function via genetic and transcriptional manipulation [10,25,74–76]. By contrast, the biochemical evaluation of their catalytic activity and substrate preferences has not received as much attention until recent years [55,77,78]. Because the biological significance of HDAs in plants has been reviewed extensively [13,62,79–81], we only include a brief summary here (Table 1).

The class I enzymes HDA1/19, HDA6, and HDA9, which have attracted most attention, are localized in the nucleus and are heavily associated with chromatin remodeling [62,80,81]. In particular, HDA6 deacetylates histones at transposable elements and helps to maintain DNA methylation at those sites, which results in gene silencing [82–84]. This function is regulated by Ser phosphorylation at the C-terminus of HDA6, which promotes its interaction with histone methyltransferases that act

<table>
<thead>
<tr>
<th>Protein</th>
<th>Localization</th>
<th>Function</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class I</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDA1/19</td>
<td>Nucleus</td>
<td>Circadian rhythm regulation, Flowering time</td>
<td>[97] [88,89,94] [87,109] [88,82,86] [44,88] [92,93,110] [22,76,88]</td>
</tr>
<tr>
<td>HDA6-1</td>
<td>Nucleus</td>
<td>Chromatin silencing, Circadian rhythm regulation, Flowering time, Phytohormone signaling, Plant growth, Seed maturation and germination, Stress response</td>
<td>[78,82–84] [85,86,90,97] [88,89,94] [88,95,111] [44,88] [98]</td>
</tr>
<tr>
<td>HDA6-2 (HDA7)</td>
<td>Nucleus</td>
<td>Plant growth, Seed maturation and germination</td>
<td>[112] [112]</td>
</tr>
<tr>
<td>HDA9</td>
<td>Nucleus</td>
<td>Flowering time, Immune response, Phytohormone signaling, Stress response, Thermomorphogenesis</td>
<td>[6,98] [102] [101,103] [96,100,101,104] [103]</td>
</tr>
<tr>
<td><strong>Class II</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDA5-1</td>
<td>Nucleus</td>
<td>Flowering time</td>
<td>[68]</td>
</tr>
<tr>
<td>HDA5-2 (HDA15)</td>
<td>Nucleus</td>
<td>Root development</td>
<td>[74]</td>
</tr>
<tr>
<td>HDA8</td>
<td>Nucleus</td>
<td>Plant growth</td>
<td>[113]</td>
</tr>
<tr>
<td>HDA14</td>
<td>Chloroplast Mitochondria</td>
<td>Microtubule dynamics, Photosynthesis regulation</td>
<td>[10] [17]</td>
</tr>
<tr>
<td>HDA15</td>
<td>Nucleus</td>
<td>Photomorphogenesis, Photosynthesis regulation, Seed germination, Stress response</td>
<td>[67] [106] [75] [76]</td>
</tr>
<tr>
<td><strong>Class IV</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDA2</td>
<td>Nucleus</td>
<td>Phytohormone signaling, Stress response</td>
<td>[114]</td>
</tr>
</tbody>
</table>
on deacetylated histones [78]. HDA6 also interacts with the histone demethylases LDL1 and LDL2 to repress the transcription of circadian rhythm-related genes [85,86]. In addition, HDA1/19 and HDA6 interact directly with a large number of transcription factors and chromatin modifiers [44,87–95], in clear analogy to the mammalian network of epigenetic deacetylase complexes [58,61,96]. In arabidopsis and other plants, HDA1/19 and HDA6 bind to the histone deacetylase complex 1 (HDC1) protein [88], which is the centerpiece for multiple protein–protein interactions with histone- and RNA-binding proteins [44,94], and to Sin3-like transcriptional regulators [94,95]. HDC1-containing complexes regulate phytohormone signaling including abscisic acid (ABA) [88,95], fruit cell proliferation [95], flowering time [94], and organ size [88]. HDA1/19 and HDA6 also form complexes with the corepressor TOPOLESS (TPL) and with pseudo-response regulators (PRRs) involved in the circadian cycle [30,97]. Interestingly, HDA1/19 forms the brassinosteroid signaling-related factor (BES1)–TPL–HDA1/19 complex that also regulates ABA signaling [32].

HDA9, although less extensively characterized, has also been proposed to form complexes with a repressive epigenetic function [98–101]. HDA9 interacts with the SANT domain-containing powdridess (PWR) and with the highly expressed osmotically responsive gene 15 (HOS15) proteins in the multifunctional HDA9–PWR–HOS15 complex [6,81,98,102]. HDA9 and PWR can also act as transcriptional activators in the context of thermomorphogenesis where they act as inducers of the synthesis of the phytohormone auxin [103]. Other epigenetic protein partners connected to HDA9 are the transcription factors VAL1 and VAL2, that are related to inducers of the synthesis of the phytohormone auxin [103]. Other epigenetic protein partners connected to HDA9 are the transcription factors VAL1 and VAL2, that are related to inducers of the synthesis of the phytohormone auxin [103].

Class II HDAs are less well understood, but they appear to fulfill highly diverse functions across several cell compartments (Table 1) [26]. HDA14 was first identified to deacetylate α-tubulin in the cytoplasm [10], similarly to HDAC6 in mammalian cells [105]. Alternatively, HDA14 has been found in mitochondria and chloroplasts [17,26] where it can regulate proteins involved in photosynthesis via deacetylation [17]. HDA15 and HDA5 are found both in the cytoplasm [26] and the nucleus, where they overlap in function with class I HDAs and modulate genes involved in the temperature stress response [76] and flowering [68]. Similarly to class I HDAs, HDA15 forms complexes with transcription factors in arabidopsis. For instance, HDA15 interacts with the phytochrome-interacting factors 1 (PIF1) and 3 (PIF3) to repress seed germination and chlorophyll biosynthesis-related genes, respectively [75,106]. HDA15 is also recruited by elongated hypocotyl 5 (HY5) to regulate light-responsive genes and photomorphogenesis [67]. Interestingly, HDA15 can shuttle between the nucleus and the cytoplasm in response to light [26]. HDA5, on the other hand, relies on phosphorylation and interaction with 14-3-3 proteins for cytoplasmic localization [26], similarly to mammalian class IIa HDACs [107,108]. HDA15 has also been reported recently to be phosphorylated at the C-terminus but, in this case, phosphorylation alters its activity and subnuclear localization [55]. Both the class II HDA8 and the class IV HDA2, that are localized in the nucleus, remain largely uncharacterized (Table 1).

Concluding Remarks and Future Perspectives

Studies over the past decade have revealed that plant HDAs play crucial roles in diverse biological processes [62]. HDAs take part in epigenetic regulatory complexes and modulate development, flowering, the circadian clock, and the response to different stresses [13,62,80,81]. However, their mechanisms of action remain largely unknown, in part owing to the lack of structural and biochemical studies. In this regard, the recent development of functional assays [12,96,115] and crystallization protocols [12,58,65,116] for mammalian HDACs could be translated to plant...
HDA and provide important insight. In this review, we provide a foundation for these advances by presenting the overall HDA phylogeny and highlighting key structural variants that warrant further investigation (see Outstanding Questions). It is our hope that the revised and updated phylogeny will also set a standard for future annotation and classification of plant HDAs. In addition, evolutionary analyses reveal, first, that Brassica species express two additional and functionally distinct HDAs, and second, that the intriguing HDA2 and HDA14 proteins retain structural characteristics from their bacterial origin. The collective findings presented in this review will underpin the elucidation of the activity and role of each individual plant HDA and, ultimately, facilitate the application of plant epigenetics to solve future challenges in ecology and crop production.

Acknowledgments
This work was supported by the Government of Aragón-European Regional Development Fund (FEDER; grant E35_17R to I. Y.).

Supplemental Information
Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tplants.2020.12.011.

References

10. Tran, H.T. et al. (2012) Arabidopsis thaliana histone deacetylase 14 (HDA14) is an α-tubulin deacetylase that associates with PPSA and enriches in the microtubule fraction with the putative histone acetyltransferase ELPS. Plant J. 71, 263–272
suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic Acids Res. 30, 5038–5055


34. Duanzhi, M.T. et al. (2011) Evolutionary origins of Brassicaceae specific genes in Arabidopsis thaliana. BMC Evol. Biol. 11, 47


40. Parrish, I.A. et al. (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. Genome Biol. 15, R77


44. Brosch, G. et al. (1995) Inhibition of male histone deacetylases by HC toxin, the host-selective toxin of Cochliobolus carbonus. Plant Cell 7, 1941–1950


64. Zhao, L. et al. (2019) HYS interacts with the histone deacetylase HDAC15 to repress hypocotyl cell elongation in photomorphogenesis. Plant Physiol. 180, 1450–1466


70. Brosch, G. et al. (1995) Inhibition of major histone deacetylases by HC toxin, the host-selective toxin of Cochliobolus carbonus. Plant Cell 7, 1941–1950


75. de Rooy, P.G.H. et al. (2020) The diverse and unanticipated roles of histone deacetylase 9 in coordinating plant development and environmental acclimation. J. Exp. Bot. 70, 6211–6225


plant stress response. Transcription factor are mutual antagonists in regulation of stress response. In a photoperiod-dependent manner. Transcriptional activator interacts with histone deacetylase. Nat. Biot. 46, 10669

modiﬁcation complex interacts with TOC1 and regulates the core circadian clock components in Arabidopsis. Front. Plant Sci. 10, 233


basal defense. LSD1 demethylase and HDAC1 deacetylase in the CoREST complex. Cell Rep. 30, 2699–2711


et al. 2011) Histone deacetylase 19 is essential for repressing circadian clock components in Arabidopsis. Science 332, 1520–1523


et al. 2006) Lenght-dependent prediction of protein intrinsic disorder. BMC Bioinforma. 7, 208


Zhang et al. (2020) Genome-wide target mapping shows histone deacetylase complex 1 regulates cell proliferation in cucumber fruit. Plant Physiol. 182, 167–184

Song et al. (2020) Mechanism of crosstalk between the LSD1 demethylase and HDA12 deacetylase in the CoREST complex. Cell Rep. 30, 2699–2711

