Supplementary figures

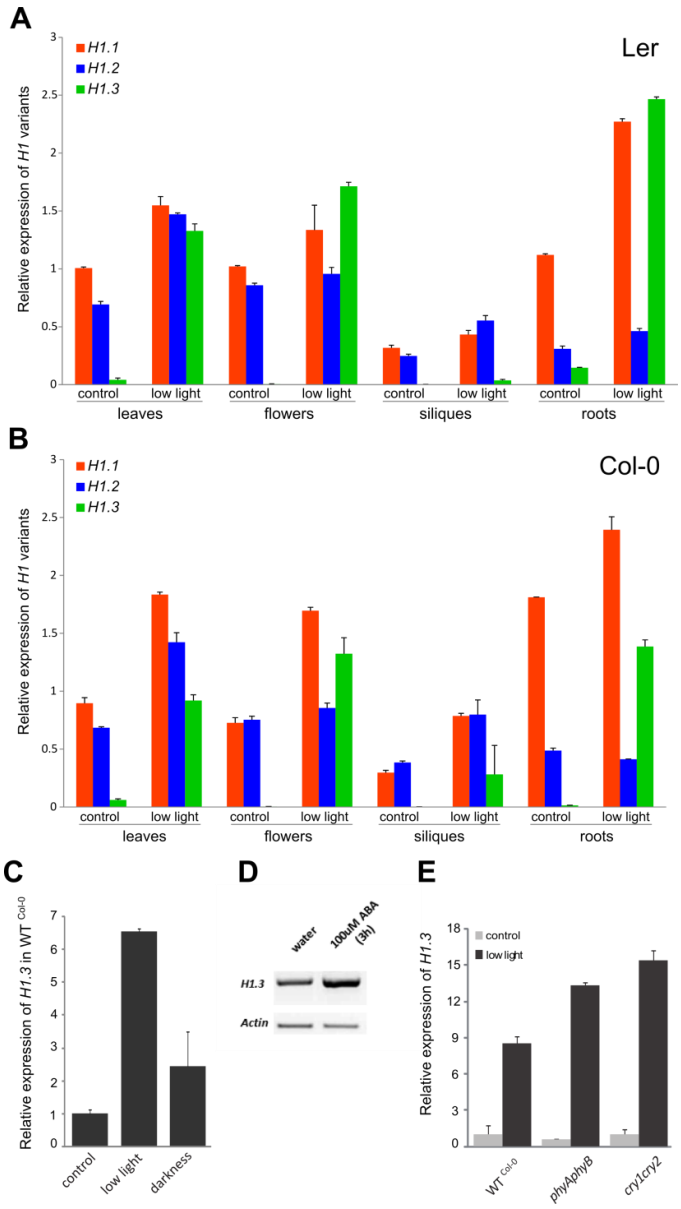

Figure S1. Characterization of *H1.3* expression. (A) Relative expression (RT-qPCR) of all *H1* variants in different tissues of Ler plants. **(B)** Relative expression (RT-qPCR) of all *H1* variants in different tissues of Col-0 plants. **(C)** Relative expression of *H1.3* in seedlings after four days of complete darkness compared with four days of limited light intensity during the day (low light) for Col-0 plants. **(D)** RT-PCR analysis of *h1.3* lines and *H1.3* transcript response to ABA treatment; *Actin* was used as a reference transcript. **(E)** Relative expression of *H1.3* in *phyAphyB* and *cry1cry2* mutants in control and 4-day low light conditions.

All qRT-PCR measurements were normalized to the expression of *UBC*. The plotted values are the means ± SD for three replicates.

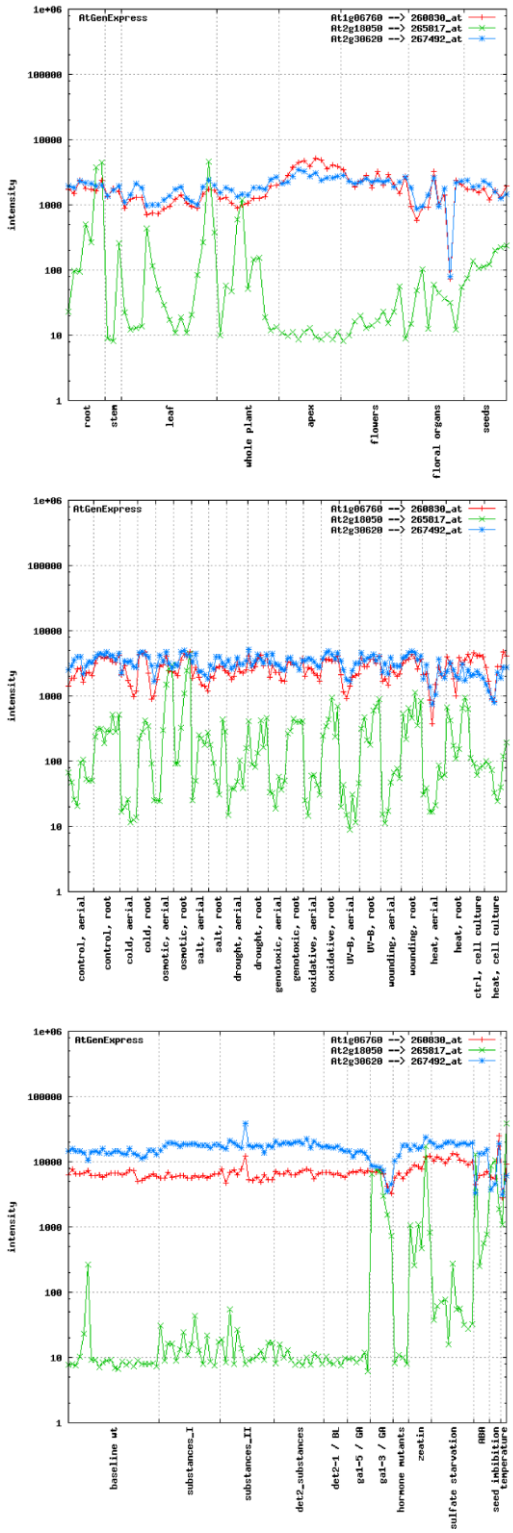

Figure S2. Expression analysis of Arabidopsis H1s using microarray data (AtGeneExpression, (Kilian et al., 2007)). The upper panel shows *H1* expression occurring in different organs. The middle panel illustrates the influence on *H1* transcription of various abiotic stresses and the bottom panel, the effect of other treatments as well as expression in mutants. Colour key: H1.1 – red; H1.2 – blue; H1.3 – green.

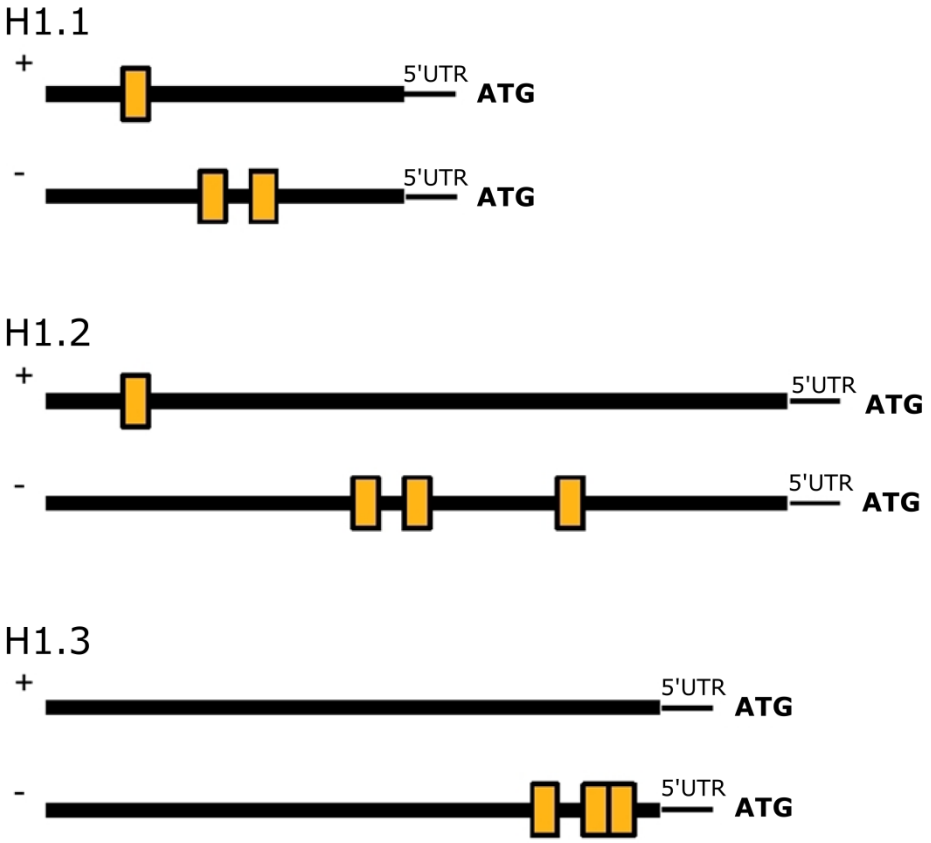

Figure S3. Schematic representation of the promoter regions of genes encoding Arabidopsis histone H1 somatic variants.

For each promoter the two complementary strands (+/-) are shown. Orange rectangles mark the location of ABRE motifs. The promoters are drawn approximately to scale. The 5'UTR sequences are not drawn to scale because no ABRE motifs were found in these regions.

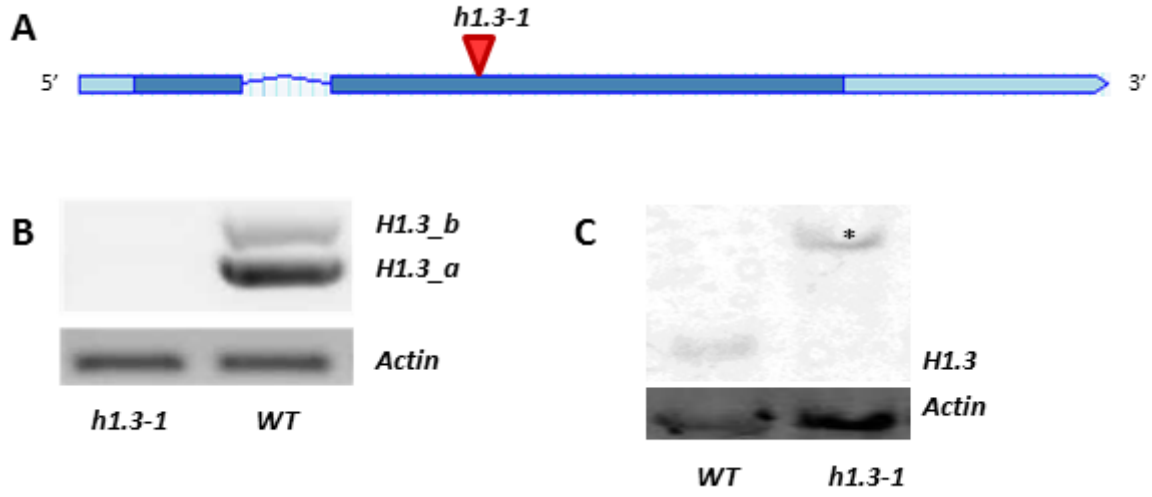

Figure S4. Characterization of the *h1.3* mutant line.

(A) Schematic diagram of the *H1.3* gene showing exons (blue bars), introns (narrow line), UTRs (light blue bars) and the transposon insertion site in the *h1.3* mutant. The H1.3 scheme was downloaded from the ABRC and modified. **(B)** RT-PCR analysis of *h1.3* lines. *H1.3_b* is a spliced form that retains an intron. At the beginning of this intron is a stop codon, so that if any protein was translated from the *H1.3_b* transcript it would only consist of part of the N-terminal domain, which is thought to have no function. Actin was used as a reference transcript. **(C)** Northern blot confirming the absence of an appropriate *H1.3* transcript in *h1.3-1*. This analysis revealed the presence of a high molecular mass band in *h1.3-1* (asterisk on panel C), which probably represents an erroneous *H1.3* transcript including part of the 6-kb transposon insertion.

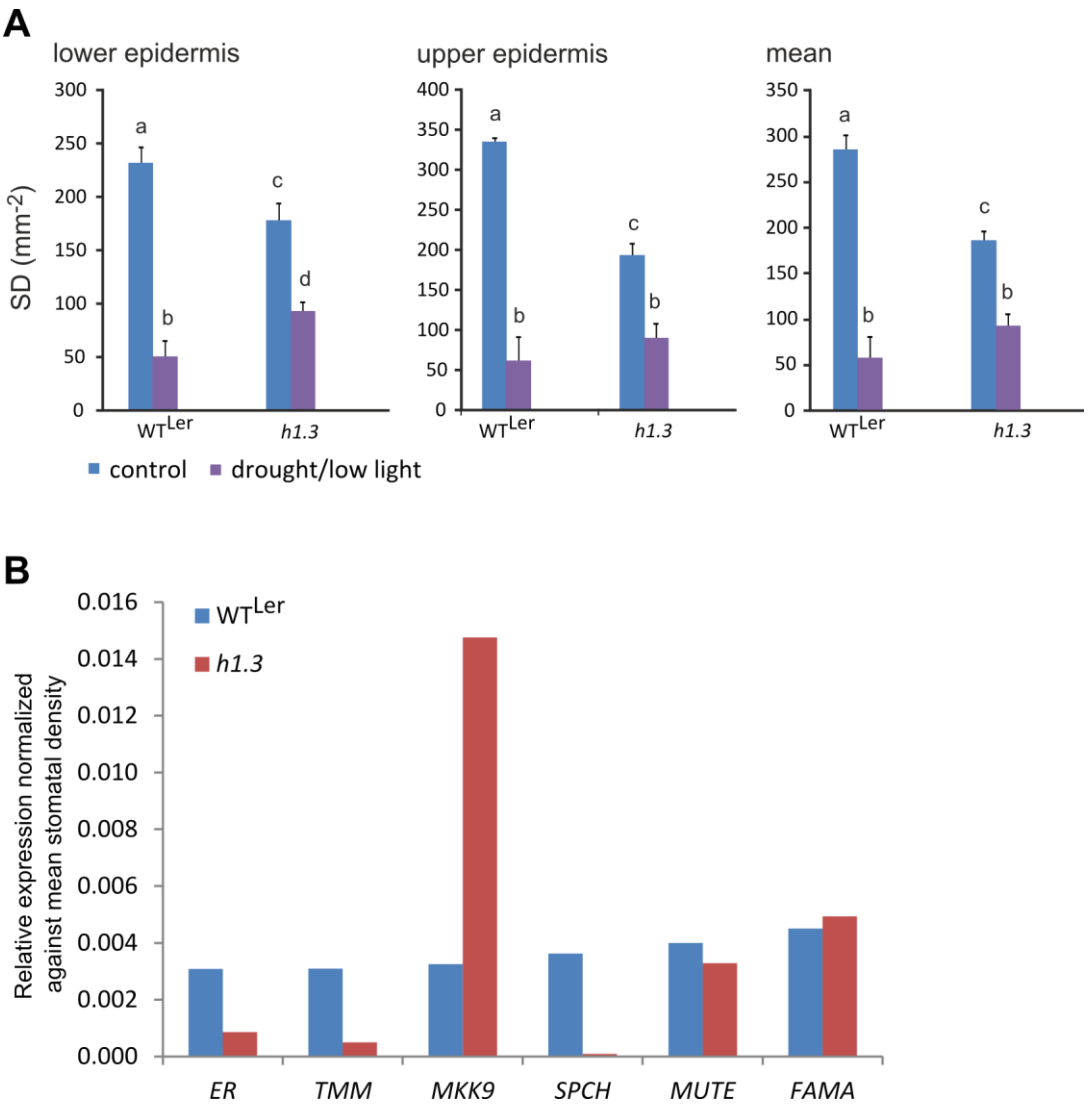

Figure S5. (A) Stomatal density (SD) in leaves of wild-type (WT^{Ler}) and *h1.3* mutant plants grown in control and drought/low light conditions. Stomatal density was measured for both the lower and upper epidermis. Statistically significant differences are indicated by letters ($p<0.05$, Tukey's HSD test). Error bars represent standard deviations. **(B)** Relative expression (RT-qPCR) of genes with key functions in guard cell biogenesis in *h1.3* mutant and wild-type plants (the mean expression value from the same data which are presented in Fig. 2C) normalized against mean stomatal density.

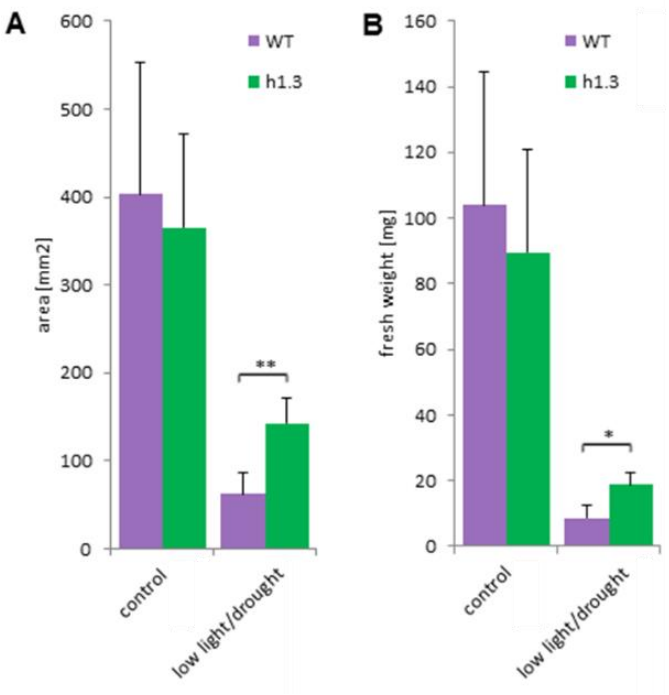

Figure S6. Growth of *h1.3* mutant plants is not restricted in response to low light/drought treatment, unlike that of wild-type plants. Color bars indicate the average area **(A)** and fresh weight **(B)** for two-week-old WT^{Ler} (purple) and *h1.3* (green) plants grown under control conditions or a combination of low light and drought. * – p-value <0.05, **– p-value ≤0.01 (T-test). The plotted values are the means ± SD.

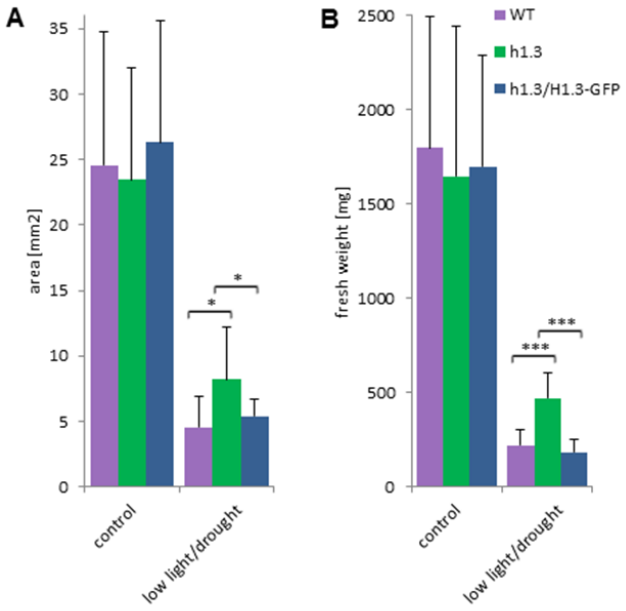

Figure S7. The *h1.3* mutant complemented with H1.3-GFP responds to combined low light/drought treatment similarly to wild-type plants.

Color bars indicate the average area **(A)** and fresh weight **(B)** for 5-week-old WT^{Ler} (purple), *h1.3* (green) and complemented *h1.3*/H1.3-GFP (dark blue) plants grown under control conditions or for 17 days with a combination of low light and drought (* – p-value <0.05, ***– p-value <0.005, T-test). The plotted values are the means ± SD.

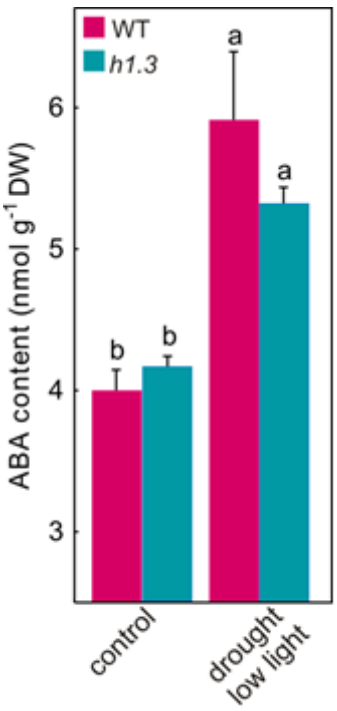

Figure S8. ABA content in wild-type (WT^{Ler}) and *h1.3* plants in control and combined low light/drought conditions. Statistically significant differences are indicated by letters ($p<0.05$, Tukey's HSD test).

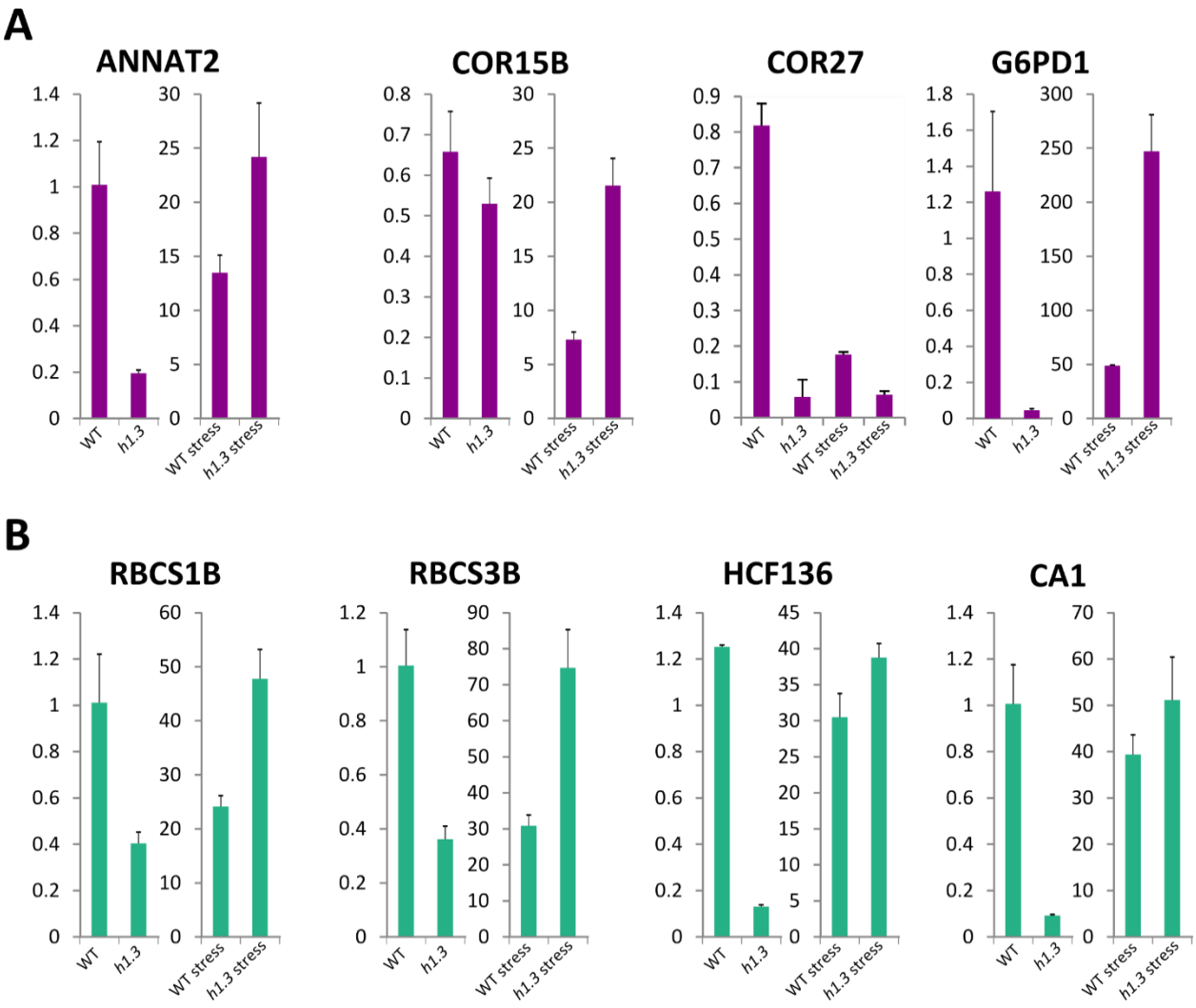

Figure S9. Verification by RT-qPCR of gene expression data obtained in microarray experiments examining the effects of combined low light/drought treatment (stress).

(A) Relative expression of genes involved in responses to environmental stimuli. **(B)** Relative expression of genes involved in responses of the photosynthetic apparatus. Error bars represent standard deviations from three replicates consisting of leaves from four plants grown in soil.

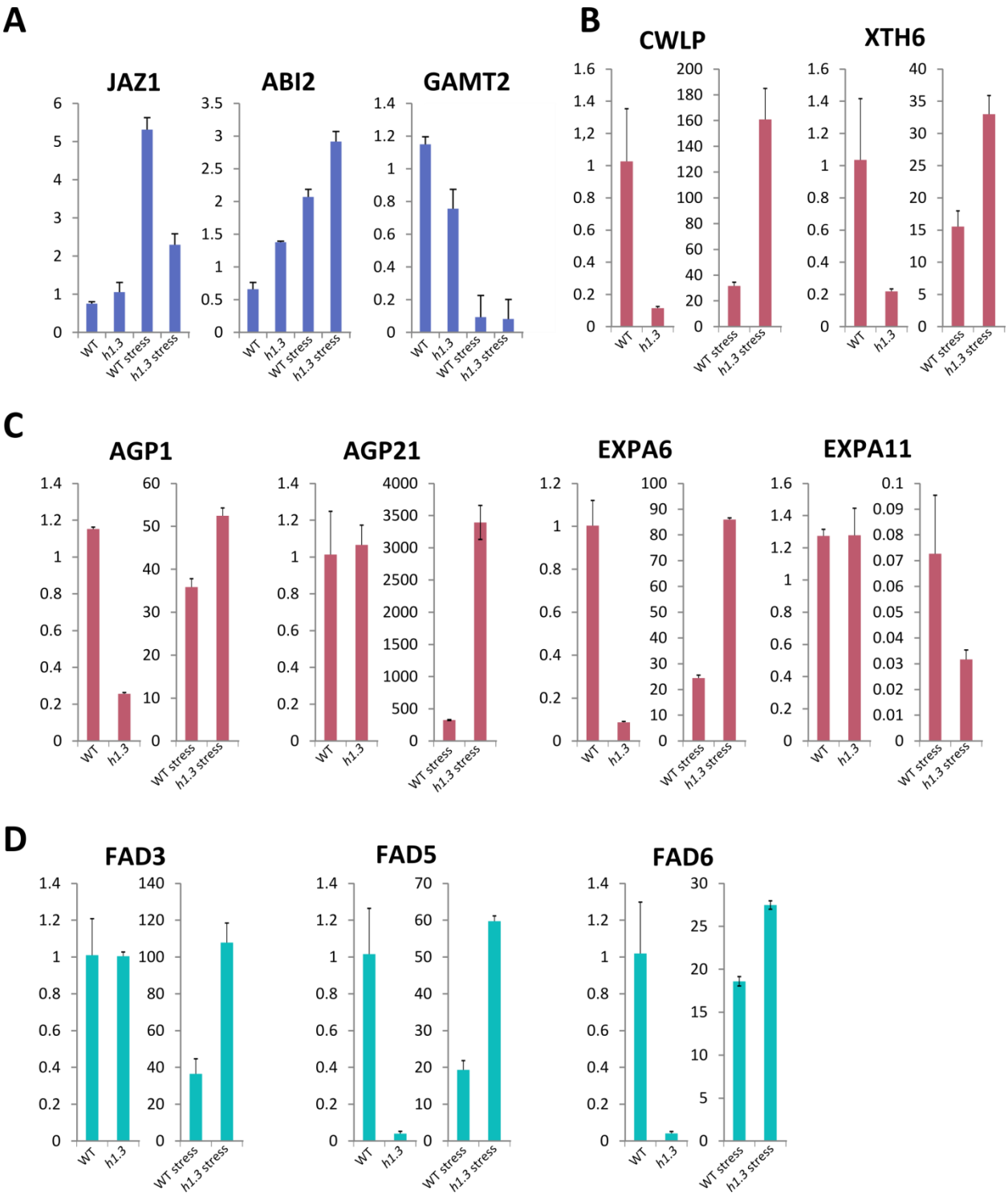

Figure S10. Verification by RT-qPCR of gene expression data obtained in microarray experiments examining the effects of combined low light/drought treatment (stress).

(A) Relative expression of genes involved in hormone regulation. **(B, C)** Relative expression of genes involved in cell wall biogenesis and regulation. **(D)** Relative expression of genes involved in fatty acid metabolism. Error bars represent standard deviations from three replicates consisting of leaves from four plants grown in soil.

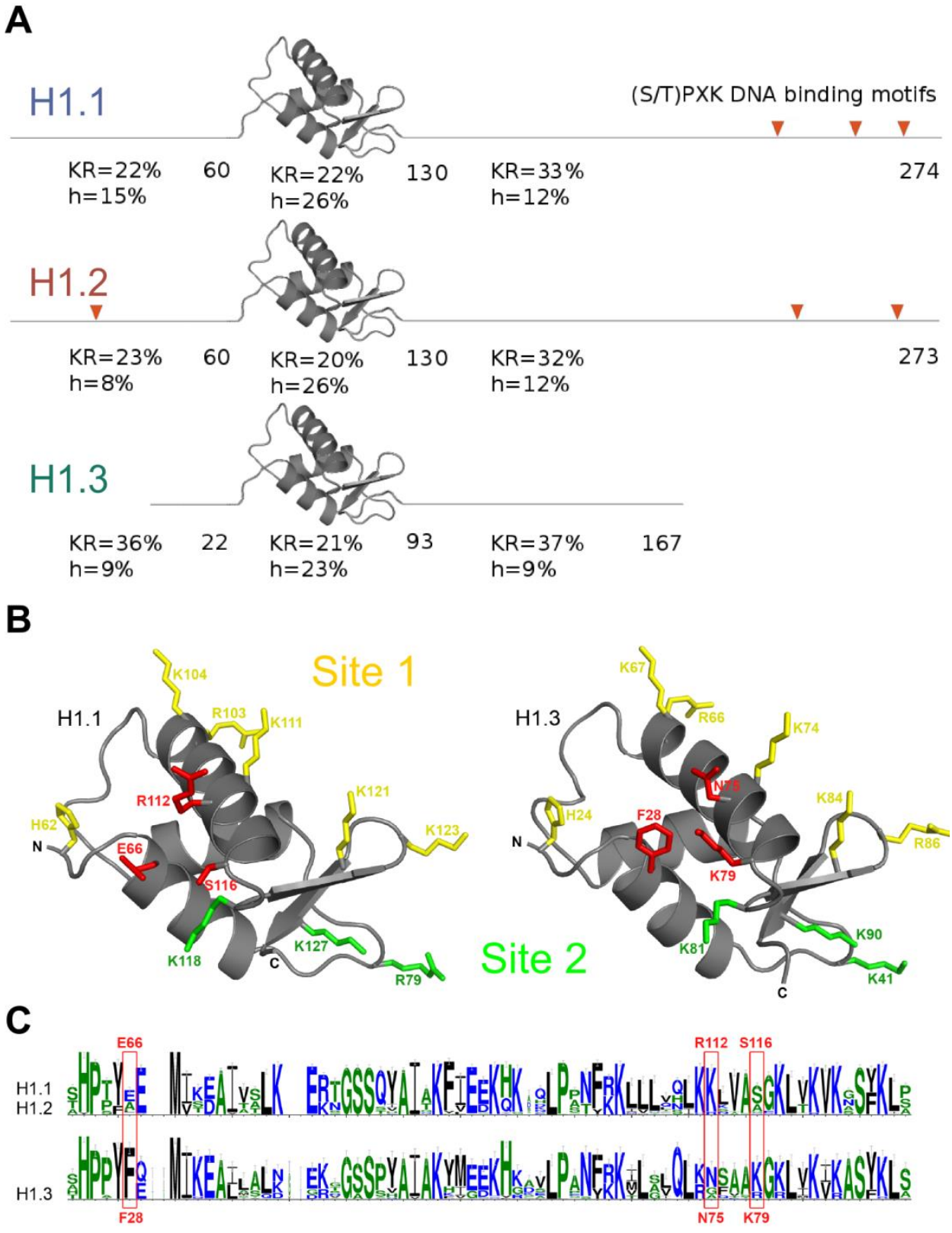
A
H1.1
(S/T)PXK DNA binding motifs
KR=22%
h=15%
60
KR=22%
h=26%
130
KR=33%
h=12%
274
H1.2
KR=23%
h=8%
60
KR=20%
h=26%
130
KR=32%
h=12%
273
H1.3
KR=36%
h=9%
22
KR=21%
h=23%
93
KR=37%
h=9%
167
B
Site 1
H1.1
K104
R103
K111
R112
K121
K123
H62
N
E66
S116
K118
K127
C
R79
H1.3
K67
R66
K74
N75
K84
H24
F28
R86
N
K79
K90
K81
K41
C
Site 2
C
E66
R112
S116
H1.1
H1.2
H1.3
F28
N75
K79

Figure S11. Arabidopsis linker histones belong to two structurally and functionally diversified families.

(A) Protein sequence features of *Arabidopsis thaliana* H1 variants. The percentages of positively charged (KR) and hydrophobic (h) residues are shown for the N-terminal tail, the central GH1 domain and the C-terminal region, respectively. Red triangles indicate (S/T)PXK motifs. **(B)** 3D models of the GH1 domain of H1.1 (left) and H1.3 (right). **(C)** Amino acid sequence conservation in the plant H1.1/2-like and H1.3-like families among flowering plant species. In the models, residues corresponding to DNA binding sites (Brown et al., 2006) are colored yellow (Site 1) and green (Site 2), while those representing the most prominent differences between plant H1.1/2-like and H1.3-like variants are shown in red**.**

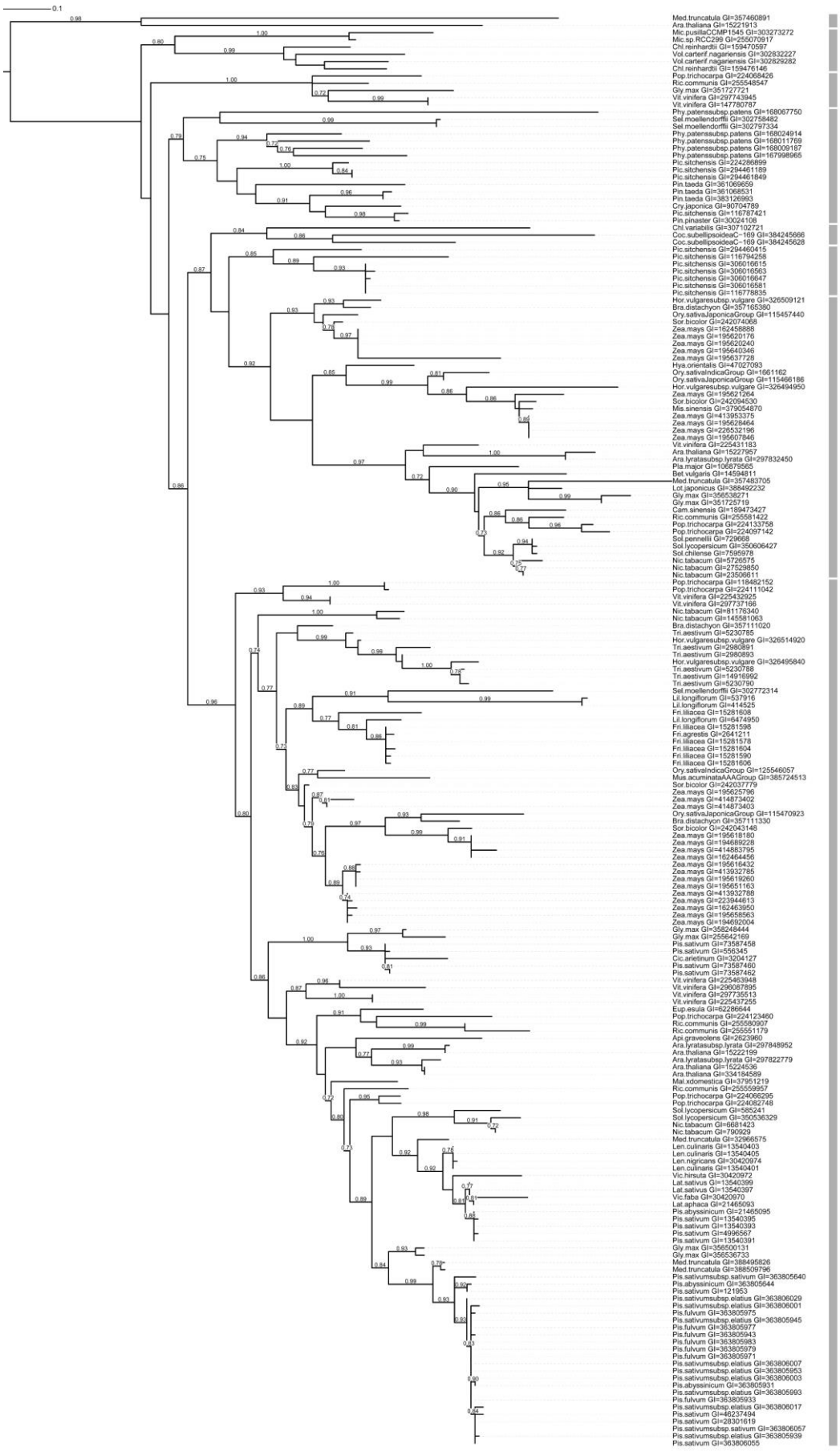
0.1
Med.truncatula GI=357460891
Ara.thaliana GI=15221913
Mic.pusillaCCMP1545 GI=303273272
Mic.sp.RCC299 GI=255070917
Chl.reinhardtii GI=159470597
Vol.carterif.nagariensis GI=302832227
Vol.carterif.nagariensis GI=302829282
Chl.reinhardtii GI=159476146
Pop.trichocarpa GI=224068426
Ric.communis GI=255548547
Gly.max GI=351727721
Vit.vinifera GI=297743945
Vit.vinifera GI=147780787
Phy.patenssubsp.patens GI=168067750
Sel.moellendorffii GI=302758482
Sel.moellendorffii GI=302797334
Phy.patenssubsp.patens GI=168024914
Phy.patenssubsp.patens GI=168011769
Phy.patenssubsp.patens GI=168009187
Phy.patenssubsp.patens GI=167998965
Pic.sitchensis GI=224286899
Pic.sitchensis GI=294461189
Pic.sitchensis GI=294461849
Pin.taeda GI=361069659
Pin.taeda GI=361068531
Pin.taeda GI=383126993
Cry.japonica GI=90704789
Pic.sitchensis GI=116787421
Pin.pinaster GI=30024108
Chl.variabilis GI=307102721
Coc.subellipsoideaC−169 GI=384245666
Coc.subellipsoideaC−169 GI=384245628
Pic.sitchensis GI=294460415
Pic.sitchensis GI=116794258
Pic.sitchensis GI=306016615
Pic.sitchensis GI=306016563
Pic.sitchensis GI=306016647
Pic.sitchensis GI=306016581
Pic.sitchensis GI=116778835
Hor.vulgaresubsp.vulgare GI=326509121
Bra.distachyon GI=357165360
Ory.sativaJaponicaGroup GI=115457440
Sor.bicolor GI=242074068
Zea.mays GI=162458888
Zea.mays GI=195620176
Zea.mays GI=195620240
Zea.mays GI=195640346
Zea.mays GI=195637728
Hya.orientalis GI=47027093
Ory.sativaIndicaGroup GI=1661162
Ory.sativaJaponicaGroup GI=115466186
Hor.vulgaresubsp.vulgare GI=326494950
Zea.mays GI=195621264
Sor.bicolor GI=242094530
Mis.sinensis GI=379054870
Zea.mays GI=413953375
Zea.mays GI=195628464
Zea.mays GI=226532196
Zea.mays GI=195607846
Vit.vinifera GI=225431183
Ara.thaliana GI=15227957
Ara.lyratasubsp.lyrata GI=297832450
Pla.major GI=106879565
Bet.vulgaris GI=14594811
Med.truncatula GI=357483705
Lot.japonicus GI=388492232
Gly.max GI=356538271
Gly.max GI=351725719
Cam.sinensis GI=189473427
Ric.communis GI=255581422
Pop.trichocarpa GI=224133758
Pop.trichocarpa GI=224097142
Sol.pennellii GI=729668
Sol.lycopersicum GI=350606427
Sol.chilense GI=7595978
Nic.tabacum GI=5726575
Nic.tabacum GI=27529850
Nic.tabacum GI=23506611
Pop.trichocarpa GI=118482152
Pop.trichocarpa GI=224111042
Vit.vinifera GI=225432925
Vit.vinifera GI=297737166
Nic.tabacum GI=81176340
Nic.tabacum GI=145581063
Bra.distachyon GI=357111020
Tri.aestivum GI=5230785
Hor.vulgaresubsp.vulgare GI=326514920
Tri.aestivum GI=2980891
Tri.aestivum GI=2980893
Hor.vulgaresubsp.vulgare GI=326495840
Tri.aestivum GI=5230788
Tri.aestivum GI=14916992
Tri.aestivum GI=5230790
Sel.moellendorffii GI=302772314
Lil.longiflorum GI=537916
Lil.longiflorum GI=414525
Fri.liliacea GI=15281608
Lil.longiflorum GI=6474950
Fri.liliacea GI=15281598
Fri.agrestis GI=2641211
Fri.liliacea GI=15281578
Fri.liliacea GI=15281604
Fri.liliacea GI=15281590
Fri.liliacea GI=15281606
Ory.sativaIndicaGroup GI=125546057
Mus.acuminataAAAGroup GI=385724513
Sor.bicolor GI=242037779
Zea.mays GI=195625796
Zea.mays GI=414873402
Zea.mays GI=414873403
Ory.sativaJaponicaGroup GI=115470923
Bra.distachyon GI=357111330
Sor.bicolor GI=242043148
Zea.mays GI=195618180
Zea.mays GI=194689228
Zea.mays GI=414883795
Zea.mays GI=162464456
Zea.mays GI=195616432
Zea.mays GI=413932785
Zea.mays GI=195619260
Zea.mays GI=195651163
Zea.mays GI=413932788
Zea.mays GI=223944613
Zea.mays GI=162463850
Zea.mays GI=195658563
Zea.mays GI=194692004
Gly.max GI=358248444
Gly.max GI=255642169
Pis.sativum GI=73587458
Pis.sativum GI=556345
Cic.arietinum GI=3204127
Pis.sativum GI=73587460
Pis.sativum GI=73587462
Vit.vinifera GI=225463948
Vit.vinifera GI=296087895
Vit.vinifera GI=297735513
Vit.vinifera GI=225437255
Eup.esula GI=62286644
Pop.trichocarpa GI=224123460
Ric.communis GI=255560907
Ric.communis GI=255551179
Api.graveolens GI=2623960
Ara.lyratasubsp.lyrata GI=297848952
Ara.thaliana GI=15222199
Ara.lyratasubsp.lyrata GI=297822779
Ara.thaliana GI=15224536
Ara.thaliana GI=334184589
Mal.xdomestica GI=37951219
Ric.communis GI=255559857
Pop.trichocarpa GI=224066295
Pop.trichocarpa GI=224082748
Sol.lycopersicum GI=585241
Sol.lycopersicum GI=350536329
Nic.tabacum GI=6681423
Nic.tabacum GI=790929
Med.truncatula GI=32966575
Len.culinaris GI=13540403
Len.culinaris GI=13540405
Len.nigricans GI=30420974
Len.culinaris GI=13540401
Vic.hirsuta GI=30420972
Lat.sativus GI=13540399
Lat.sativus GI=13540397
Vic.faba GI=30420970
Lat.aphaca GI=21465093
Pis.abyssinicum GI=21465095
Pis.sativum GI=13540395
Pis.sativum GI=13540393
Pis.sativum GI=4996567
Pis.sativum GI=13540391
Gly.max GI=356500131
Gly.max GI=356536733
Med.truncatula GI=388495826
Med.truncatula GI=388509796
Pis.sativumsubsp.sativum GI=363805640
Pis.abyssinicum GI=363805644
Pis.sativum GI=121953
Pis.sativumsubsp.elatius GI=363806029
Pis.sativumsubsp.elatius GI=363806001
Pis.fulvum GI=363805975
Pis.sativumsubsp.elatius GI=363805945
Pis.fulvum GI=363805977
Pis.fulvum GI=363805943
Pis.fulvum GI=363805983
Pis.fulvum GI=363805979
Pis.fulvum GI=363805971
Pis.sativumsubsp.elatius GI=363806007
Pis.sativumsubsp.elatius GI=363805953
Pis.sativumsubsp.elatius GI=363806003
Pis.abyssinicum GI=363805931
Pis.sativumsubsp.elatius GI=363805993
Pis.fulvum GI=363805933
Pis.sativumsubsp.elatius GI=363806017
Pis.sativum GI=46237494
Pis.sativum GI=28301619
Pis.sativumsubsp.sativum GI=363806057
Pis.sativumsubsp.elatius GI=363805939
Pis.sativum GI=363806055
HMG
Chlorophyta
Magnoliophyta
Bryophyta
Lycopodiophyta
Pinophyta
Chlorophyta
Pinophyta
Magnoliophyta
H1.3
Magnoliophyta
H1.1-2

Figure S12. Phylogenetic tree of 196 plant H1 proteins with HMG sequences used as an outgroup.

The phylogenetic analysis was performed with PhyML. Approximate likelihood ratio test SH-like branch supports of above 50% are shown. The tree images were prepared with iTol. Phylogenetic trees inferred using different sequence datasets, outgroups and parameter settings produce variable topologies with regard to the old *Viridiplantae* clades. The long branches belonging to moss and algae can be positioned either within angiosperm clades or at the base of the tree. This uncertain positioning is possibly due to a combination of two factors: the limited number of phylogenetically informative sites and the phenomenon of long-branch attraction. The H1.1 and H1.3 clades are reproducible. Recurrent Whole Genome Duplications and polyploidization events have been vital to plant genome evolution. Multiple duplications and gene losses shaped the evolutionary history of the H1 protein family; in consequence there are multiple in- and out-paralogs, e.g. in *Zea mays* and *Pisum sativum*, which are visible as species-specific or order-specific sequence clusters.


Amoebozoa
Chlorophyta
Lycopodiophyta
Magnoliophyta
Bryophyta
Lycopodiophyta
Pinophyta
Pinophyta
Chlorophyta
Magnoliophyta
H1.3
Magnoliophyta
H1.1-2

Figure S13. Phylogenetic tree of 274 plant H1 proteins with *Dictyostelium discoideum* H1 protein as an outgroup. See the detailed description under Figure S12.

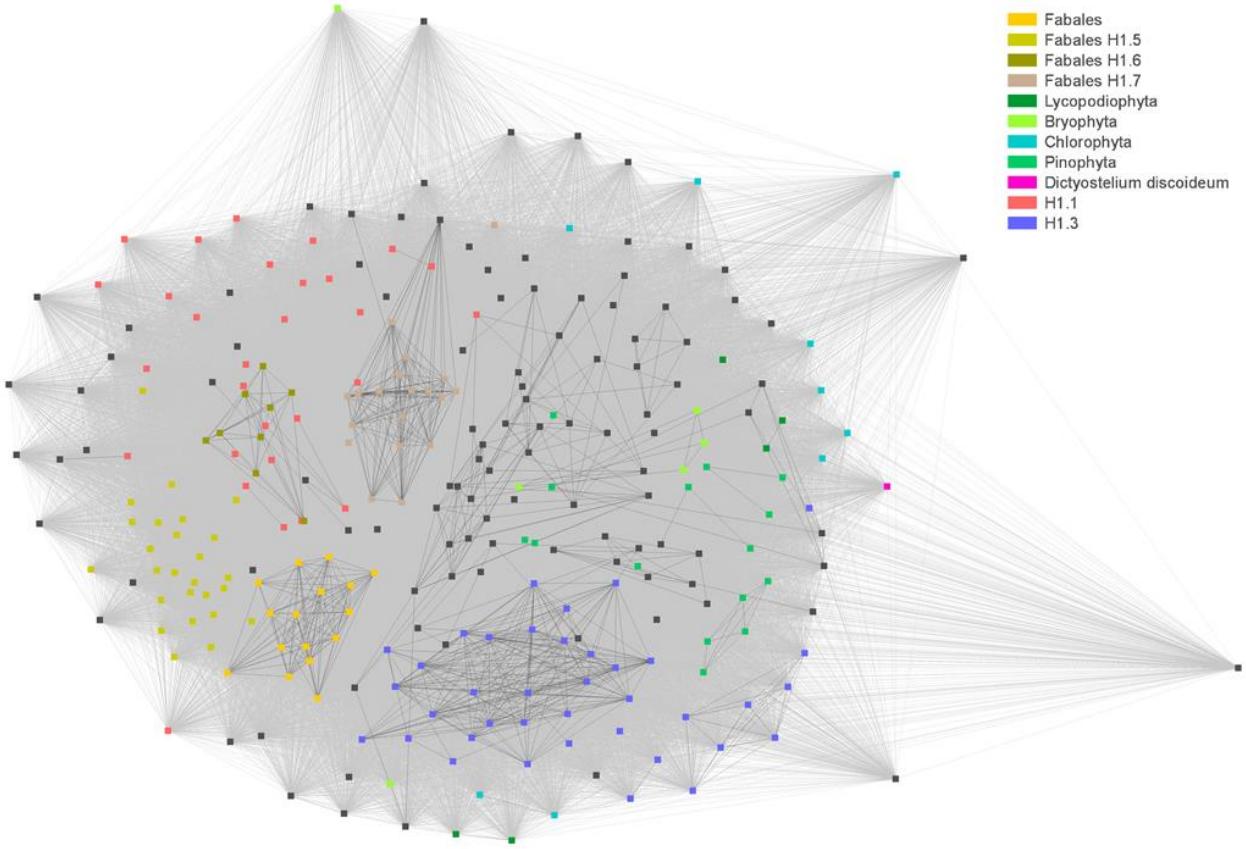

Figure S14. CLANS clustering of 274 plant H1 proteins and a 2D image of the clustering results for an interactive graphical representation of *Viridiplantae* H1 sequences.

To elucidate the relationships between and within a group of sequences we used CLANS, a graphical clustering tool implemented in Java, which visualizes pairwise sequence similarities. Sequences in the graph are represented as vertices and all-against-all BLAST high-scoring segment pairs (HSPs) as edges. The H1.1 representatives form species-specific clusters, e.g. for Fabales. H1.3 sequences are mostly grouped in one part of the diagram. Algae, fern, moss and gymnosperm sequences neither form a specific cluster nor enter one of the aforementioned groups. These general observations are consistent with the phylogenetic analyses. It should be emphasized that clustering is alignment independent and can be applied to large data sets. The H1 proteins are extremely ancient and their sequences are fairly well conserved between distantly related taxa. For example, the *Chlorella* H1 sequence has a similar level of amino acid sequence identity (60%) to both angiosperms and *Micromonas*. In addition, the GH1 domain of the slime mold *Dictyostelium* displays 50% sequence identity to that of plants. The GH1 protein domain is only about 70 amino-acids long and the evolutionary distance between the aforementioned lineages is far greater than that between

major plant groups that evolved after land colonization. Given this high level of sequence conservation over a very long evolutionary time, one might hypothesize that since functional GH1 protein domains have had to retain almost half of their residues, in consequence their sequence similarity cannot fall below a certain level. The observed similarity might be the result of substitution saturation within the limited functional space.

For a more reliable classification and in order to study the taxonomic distribution of both H1 variants, we performed detailed phylogenetic analyses. Clades grouping sequences of the main (H1.1/2-like) and stress-inducible (H1.3-like) variants from all mono- and dicotyledonous plants (for which the relevant information is available) are clearly separated. Phylogenetic inference demonstrated that H1-type globular domains from moss, fern or gymnosperms cannot be precisely classified into main and stress-inducible variants. Their basal positioning on the phylogenetic tree suggests that divergence of the two clades of angiosperm H1s was a later evolutionary event, probably associated with the divergence of angiosperms. *Amborella trichopoda*, a member of the Amborellales, the earliest angiosperm branch (Magnoliophyta), already has both H1.1/2-like (gi|548841348) and H1.3-like (gi|548856617) GH1 variants that also display the characteristic differences in CTD length.

A temporary score file is presented, with the run file available online:
http://bioputer.mimuw.edu.pl/data/H1/CLANS

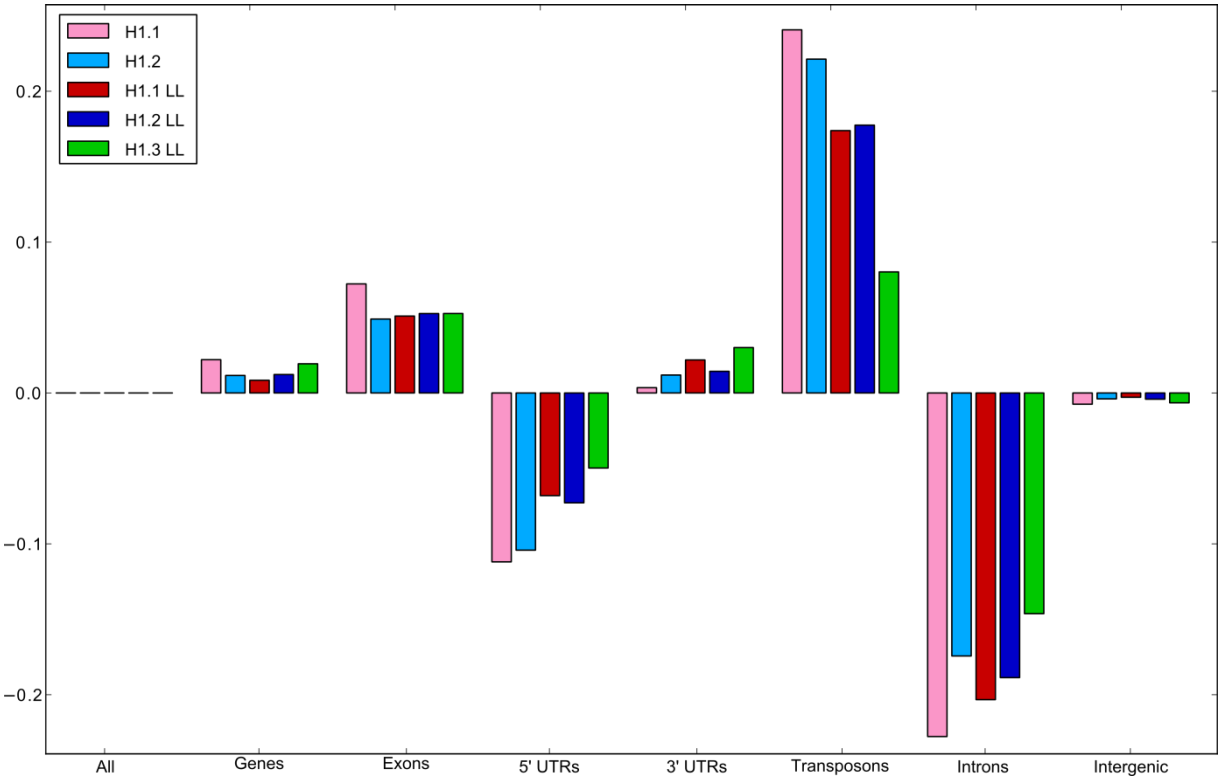

Figure S15. Genomic profiles for H1 variants in plants grown in control and 4-day low light (LL) conditions. H1 variants are depleted on 5'UTRs and introns, and overrepresented on transposons. The length of the colored bars represents the level of enrichment or depletion of the H1 variant compared with the total genome signal (All). There are no statistically significant differences for the H1.1 and H1.2 variants. The complete statistics are presented in Table S3.

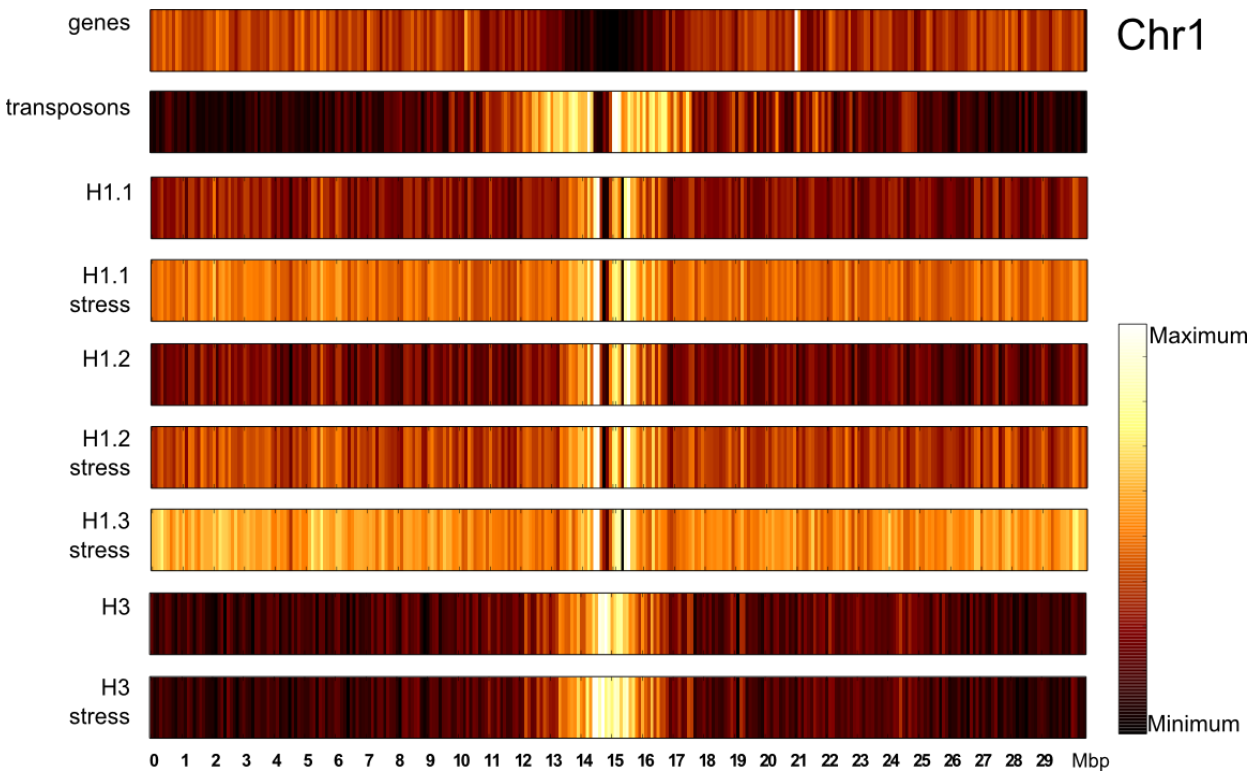

Figure S16. Distribution of H1 variants along Arabidopsis chromosome 1 in control and low light conditions. The heatmaps show the distribution of the H1 signal (ln(IP/Input)) along the whole chromosome (averaged for each 100 kb). Regions with lighter colors indicate a stronger signal compared to darker regions within one heatmap bar. For comparison, the distribution of genes and transposons is shown at the top. The unit of measurement on the bottom scale is 1 Mbp. It is important to note that since ChIP-on-chip data do not allow quantitative estimation of absolute protein amounts, only qualitative distribution profiles can be compared between variants. The heatmap scale can however be used to analyze quantitative differences in occupancy for a certain variant in particular conditions.

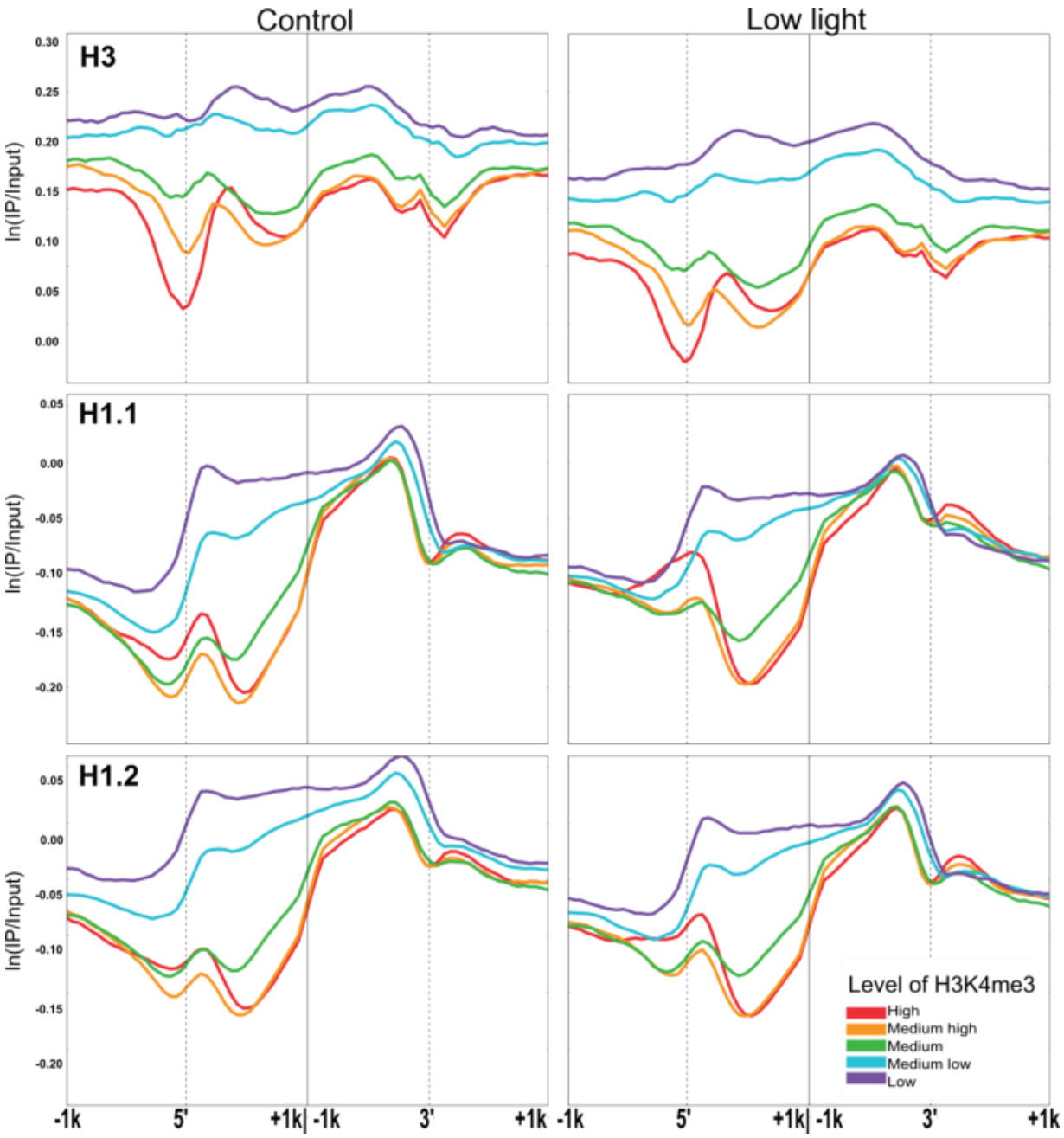

Figure S17. Distribution of the main H1s and H3 within genes with different levels of H3K4me3 of plants grown in control and 4-day low light conditions. Genes are divided into five groups according to the level of the H3K4me3 mark (Luo et al., 2012). The signals for occupancy by H1s and H3 are plotted for 1 kb around both the 5' (TSS) and 3' (TTS) ends for the five classes of genes. Each group consists of at least 4495 genes.

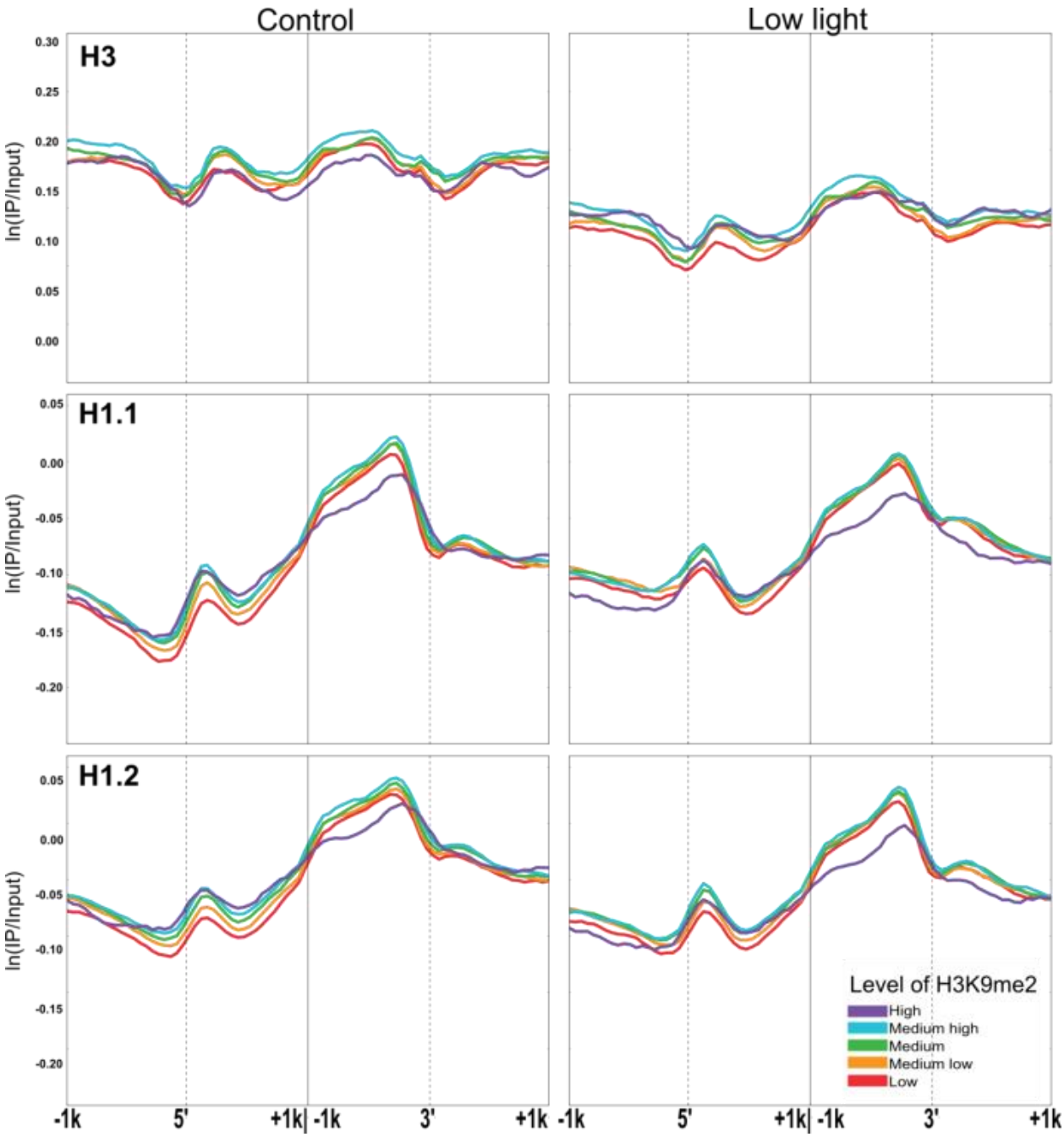

Figure S18. Distribution of the main H1s and H3 within genes with different levels of H3K9me2 of plants grown in control and 4-day low light conditions. Genes are divided into five groups according to the level of the H3K9me2 mark (Moissiard et al., 2012). The signals for occupancy by H1s and H3 are plotted for 1 kb around both 5' (TSS) and 3' (TTS) ends for the five classes of genes. Each group consists of at least 4495 genes.

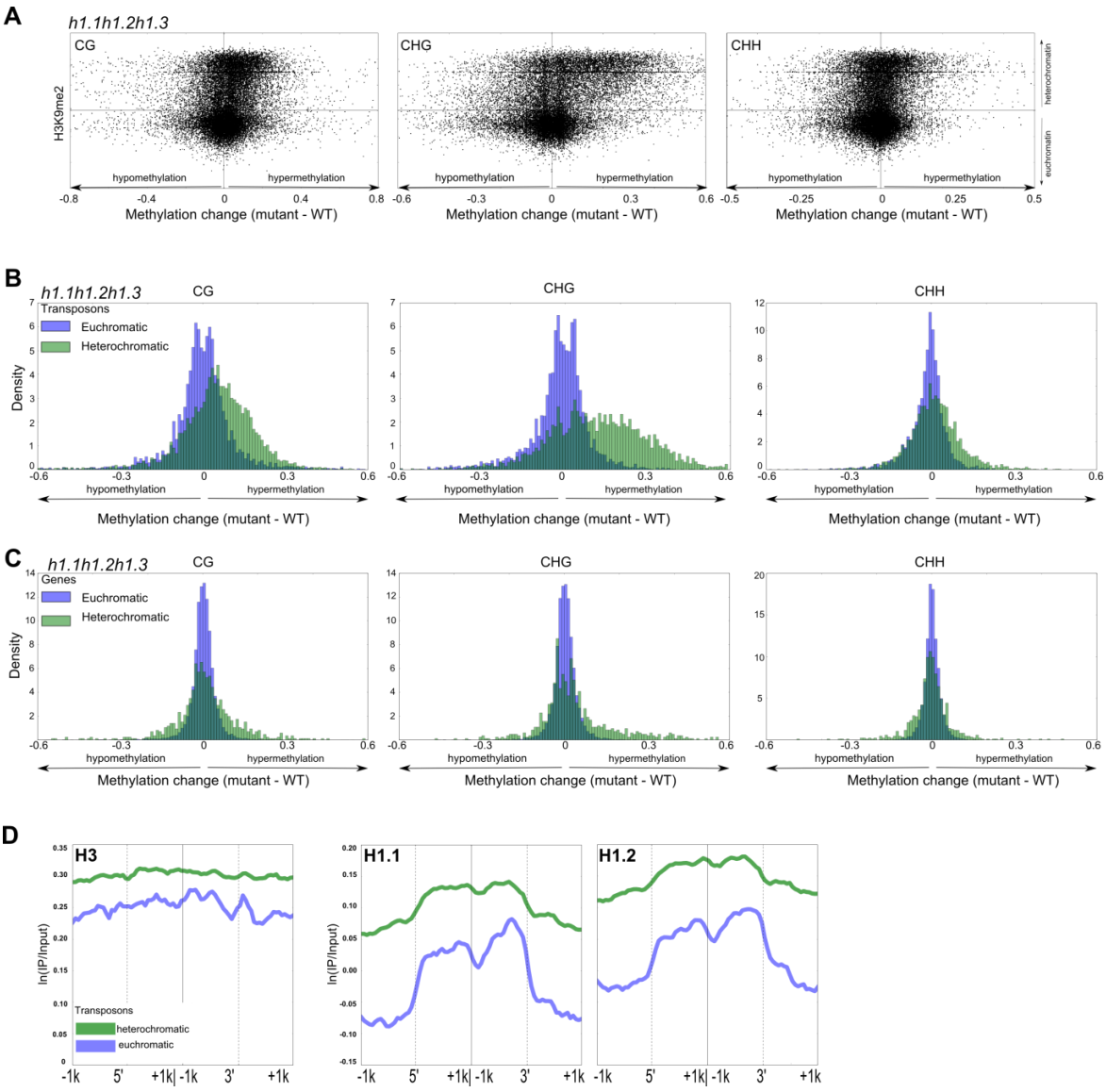
A
h1.1h1.2h1.3
CG
CHG
CHH
H3K9me2
heterochromatin
euchromatin
hypomethylation
hypermethylation
Methylation change (mutant - WT)
B
h1.1h1.2h1.3
Transposons
Euchromatic
Heterochromatic
Density
C
Genes
D
H3
H1.1
H1.2
ln(IP/Input)
heterochromatic
euchromatic
-1k
5'
+1k
3'

Figure S19. Methylation levels in the triple *h1.1h1.2h1.3* mutant. (A) Density plots of differences in the methylation of transposable elements (TEs) in *h1.1h1.2h1.3* and wild-type (WT^{Col-0}) plants. Positive numbers indicate higher methylation levels in the triple *h1* mutant. The higher the level of H3K9me2, the more heterochromatic the location of the TE. **(B)** Kernel density plots of methylation differences between *h1.1h1.2h1.3* and wild-type plants for euchromatic and heterochromatic transposons. For euchromatic transposons, the quintal with the lowest H3K9me2 and for heterochromatic transposons, the quintal with the highest H3K9me2, were selected from all transposons. **(C)** Kernel density plots of methylation differences between *h1.1h1.2h1.3* and wild-type plants on euchromatic and heterochromatic genes. Euchromatic genes were selected as those with a low level of H3K9me2, and heterochromatic genes as those with a high level of H3K9me2. **(D)** Distribution of the main H1 variants and H3 in control conditions across transposons divided between heterochromatin and euchromatin. The average distribution of H1s around the 3' and 5' ends (+/- 1kb) of TEs is plotted. TEs are divided into euchromatic TEs and heterochromatic TEs, according to their level of H3K9me2 occupancy.

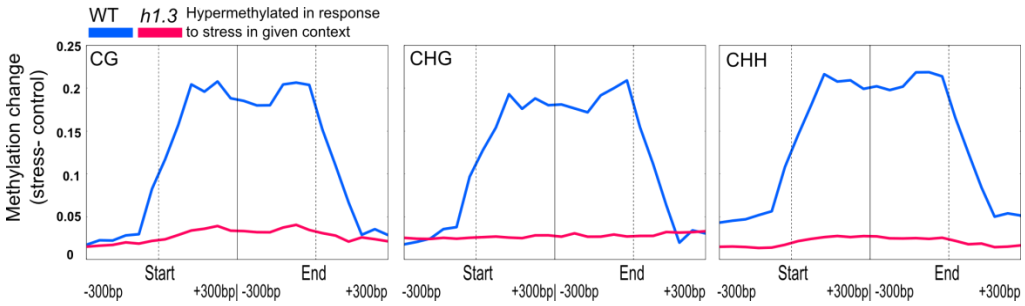

Figure S20. Methylation changes in response to stress in wild-type (WT^{Ler}) and *h1.3* mutant plants. Complementary to Figure 8B.

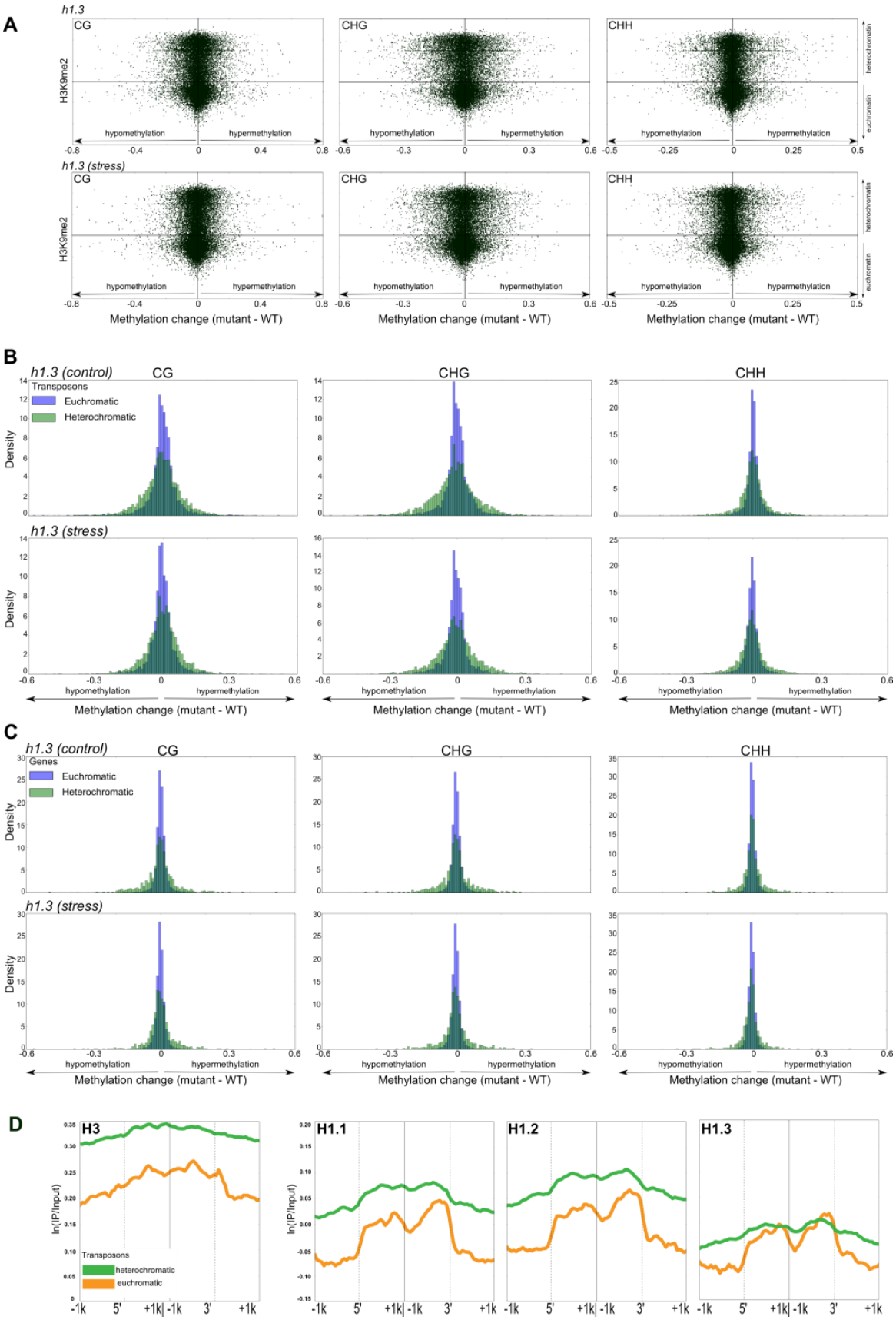
A
h1.3
CG
CHG
CHH
H3K9me2
heterochromatin
euchromatin
hypomethylation
hypermethylation
h1.3 (stress)
Methylation change (mutant - WT)
B
h1.3 (control)
Transposons
Euchromatic
Heterochromatic
Density
h1.3 (stress)
C
Genes
D
H3
H1.1
H1.2
H1.3
ln(IP/Input)
Transposons
heterochromatic
euchromatic
-1k
5'
+1k
3'

Figure S21. Global methylation changes in the *h1.3* mutant.

(A) Density plots of differences in the methylation of euchromatic and heterochromatic transposons in *h1.3* and wild-type plants grown in control or combined low light/drought conditions. For euchromatic transposons, the quintal with the lowest H3K9me2 and for heterochromatic transposons, the quintal with the highest H3K9me2, were selected from all transposons. **(B)** Kernel density plots of differences in the methylation of euchromatic and heterochromatic transposons in *h1.3* and wild-type plants grown in control or combined low light/drought conditions. For euchromatic transposons, the quintal with the lowest H3K9me2 and for heterochromatic transposons, the quintal with the highest H3K9me2, were selected from all transposons. **(C)** Kernel density plots of differences in the methylation of euchromatic and heterochromatic genes in *h1.3* and wild-type plants grown in control or combined low light/drought conditions. Euchromatic genes were selected as those with low levels of H3K9me2 and heterochromatic genes as those with high levels of H3K9me2. **(D)** Distribution of the main H1 variants, H1.3 and H3 in stress conditions across transposons divided between heterochromatin and euchromatin. The average distribution of H1s is plotted around the 3' and 5' ends (+/- 1kb) of TEs. TEs are divided into euchromatic TEs and heterochromatic TEs, according to their level of H3K9me2 occupancy.

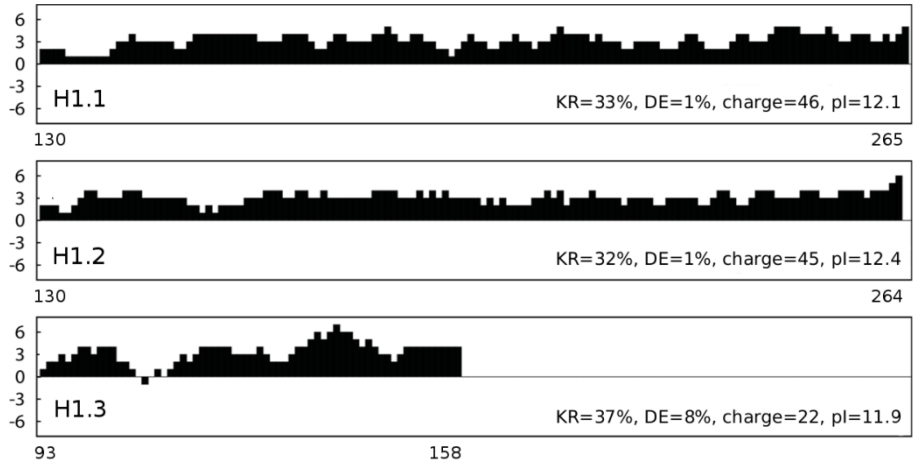

Figure S22. Moving sum plot of net charge for the C-terminal region of Arabidopsis histone H1 variants.

The net charge (y-axis) is summed in a 10-aa sliding window, with the position along the CTD denoted on the x-axis. For each CTD, 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. The average charge per unit length is similar for all variants (the same is true for all analyzed species).