Current Biology, Volume *30*

Supplemental Information

Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin

Sean A. Montgomery, Yasuhiro Tanizawa, Bence Galik, Nan Wang, Tasuku Ito, Takako Mochizuki, Svetlana Akimcheva, John L. Bowman, Valérie Cognat, Laurence Maréchal-Drouard, Heinz Ekker, Syuan-Fei Hong, Takayuki Kohchi, Shih-Shun Lin, Li-Yu Daisy Liu, Yasukazu Nakamura, Lia R. Valeeva, Eugene V. Shakirov, Dorothy E. Shippen, Wei-Lun Wei, Masaru Yagura, Shohei Yamaoka, Katsuyuki T. Yamato, Chang Liu, and Frédéric Berger

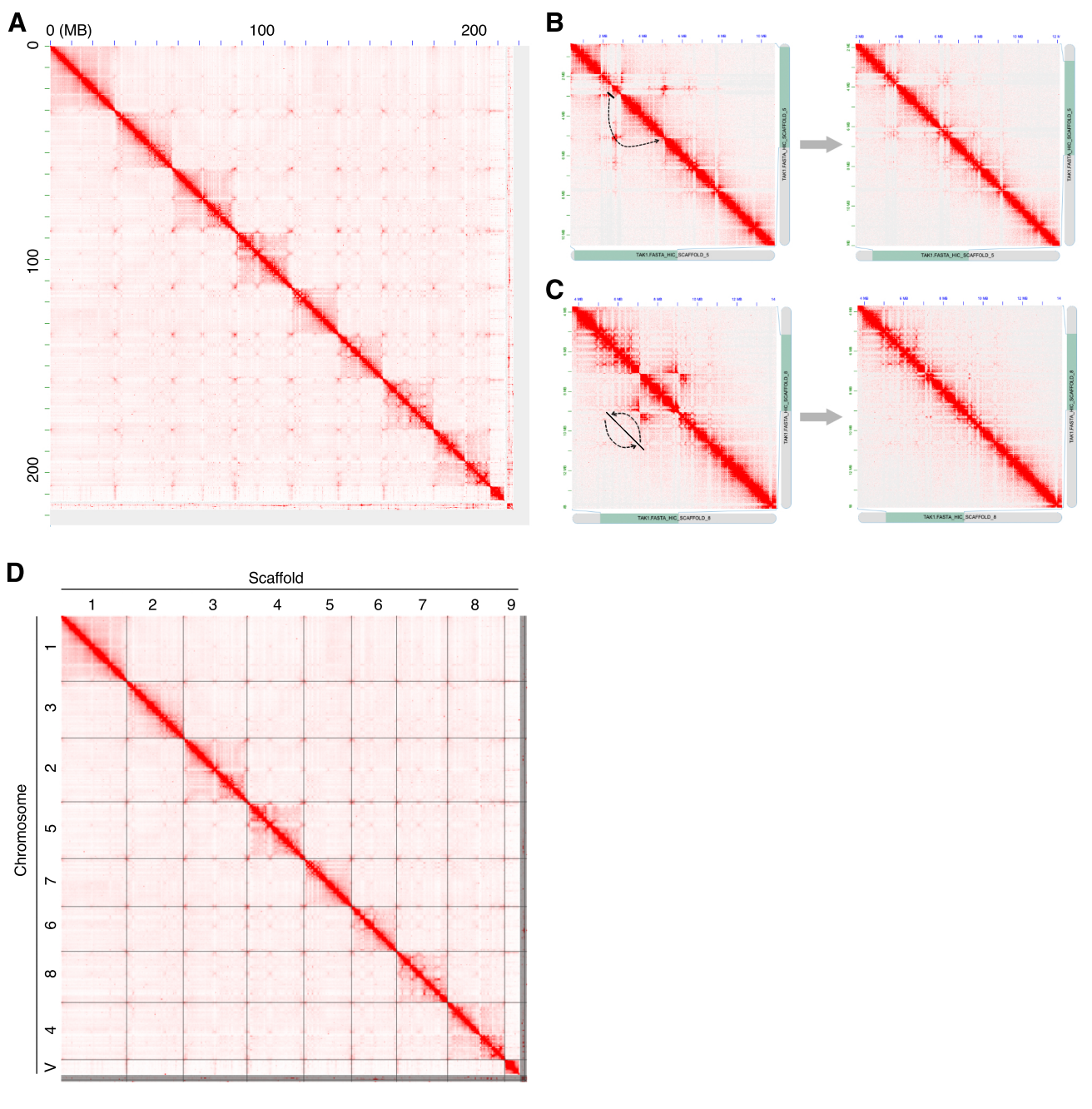
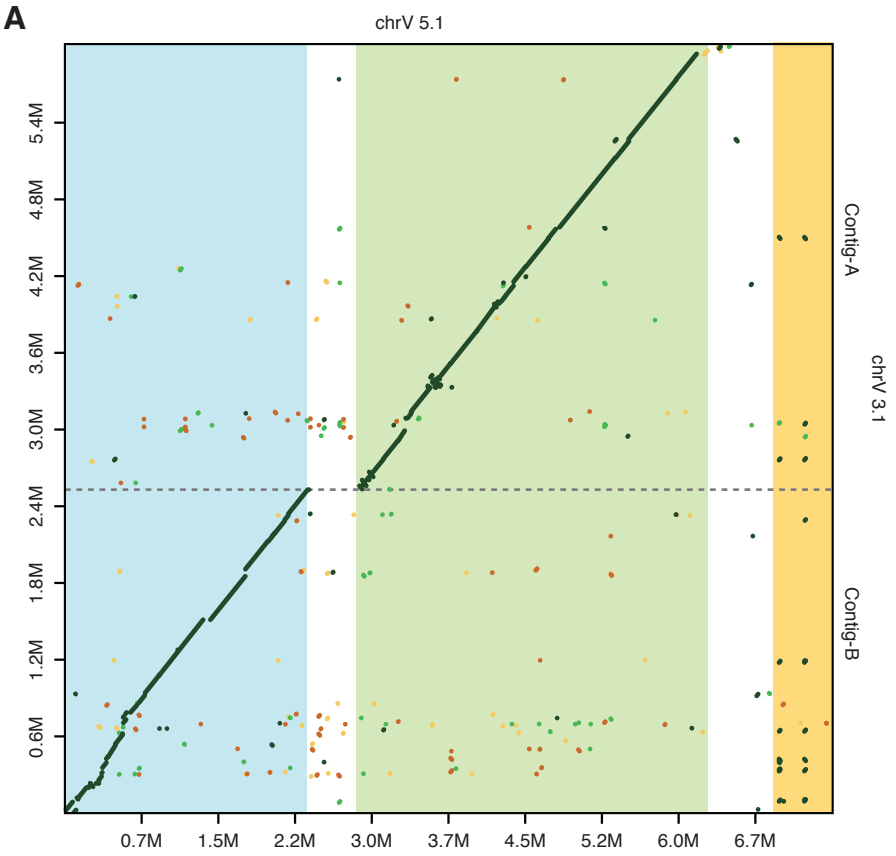

A
0 (MB)
100
200
0
100
200
B
C
D
Scaffold
1
2
3
4
5
6
7
8
9
Chromosome
1
3
2
5
7
6
8
4
V

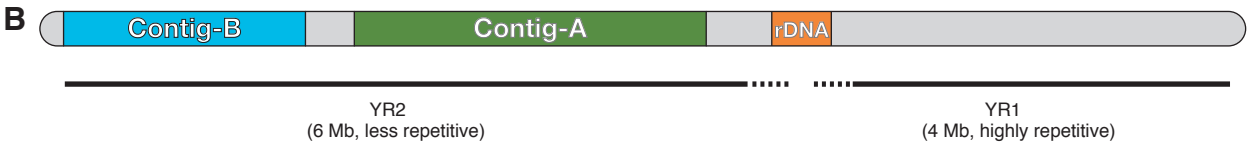
Figure S1. Hi-C guided assembly of the *Marchantia* Tak-1 genome. Related to STAR Methods.

(A) Hi-C map of the assembled "super-scaffold", visualized with Juicebox [S1]. The vast majority of this super-scaffold is consisting of 9 distinct self-interacting blocks.

(B and C) Manual inspection and correction of local misjoints. Panels on right show corrected Hi-C maps. Depending on the nature of aberrant interaction patterns, they can be corrected by changing the order of scaffolds, such as shift (B), inversion (C), or a combination of them.

(D) Hi-C map of the Tak-1 genome with manual correction. The nomenclature of chromosomes 1 to 8 and chromosome V is according to the sizes of the longest assembled 9 scaffolds.





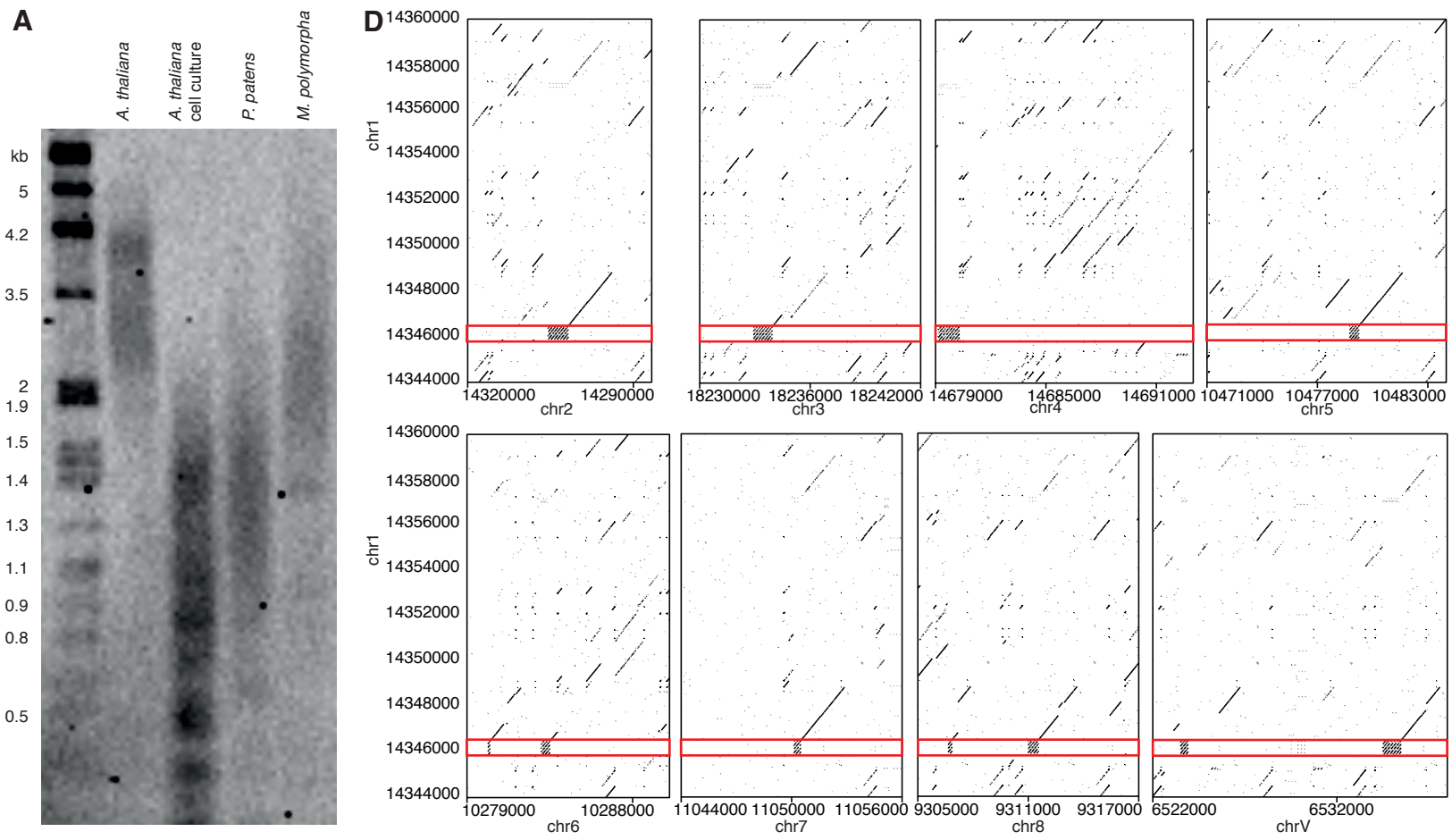
YR2
(6 Mb, less repetitive)

YR1
(4 Mb, highly repetitive)

Figure S2. Chromosome V structure. Related to Figure 5.

(A) Comparison of the chromosome V sequences from the genome assembly versions 5.1 (this study) and 3.1[S2]. The regions corresponding to those previously sequenced (Contig-A and Contig-B; [S3]) and an rDNA cluster are colored. The rDNA cluster contains 6 copies of rDNA repeat unit, which shows 99.6% and 97.0% similarities to the autosomal and U-chromosomal copies[S4], respectively.

(B) Schematic diagram of the V chromosome structure. The V-chromosomal segments, YR1 and YR2, identified in the previous study [S3] are represented by thick lines. Note that the boundary between YR1 and YR2 is not determined.



B Centromeric repeat

Probe 1

TGTTGGAGTGTCAACAGGAGCTGAGATTTAGTGGGCTTGTTCACGACGGCCGGG
CGCACATACCTGCAAATTTTCAGCCCCAACGGAGCTGCTGTCAAGAAGTTGTCAT
TTCGAAACTTTGAGTTTAAGGGTTTGAGCATGAATCTTGCGGGGGCCATATCTTCT

Probe 2

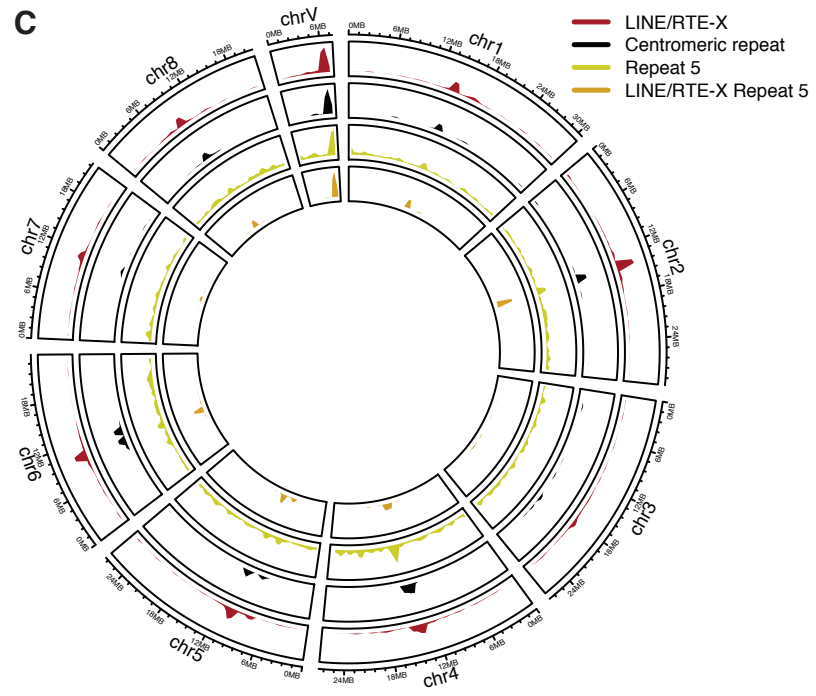


Figure S3. Centromeres and telomeres in *Marchantia*. Related to Figure 1.

(A) Comparative TRF (telomere repeat fragment assay, Southern blotting) analysis of DNA isolated from *Arabidopsis thaliana* Col-0 ecotype, *Arabidopsis thaliana* cell culture, *Physcomitrella patens* strain Gransden and *Marchantia polymorpha*. As expected, telomeres in all plants display a heterogeneous profile, but the mean length differs between the species. *M. polymorpha* telomeres (mean TRF 2,058 bp) are shorter than in *A. thaliana* (mean TRF 2,976 bp), but longer than in the model moss *P. patens* (mean TRF 1,443 bp). Telomere lengths of *A. thaliana* cell culture are shorter than in plants. The blot was hybridized with a DIG-labeled $(TTTAGGG)_4$ probe. Molecular weight markers are shown on the left.

(B) Centromeric repeat sequence. Probes used for FISH experiments are highlighted.

(C) Circos plot of centromere-related feature distributions across the genome. Each band shows the density of each feature per chromosome, relative to the greatest density per band. Centromeric repeat band based on positions of BLAST hits with E-values < 10^{-30} using the putative centromeric as a query against the *Marchantia* genome. LINE/RTE-X Repeat 5 band corresponds to all LINE/RTE-X elements belonging to repeat cluster 5.

(D) Dot plots between chromosome 1 and each other chromosome. Putative centromeric repeat is highlighted on chromosome 1. Sequences with a score greater than 50 over 10bp windows appear as dots. Genomic coordinates shown along the x-axis and are the same for chromosome 1 in each plot.

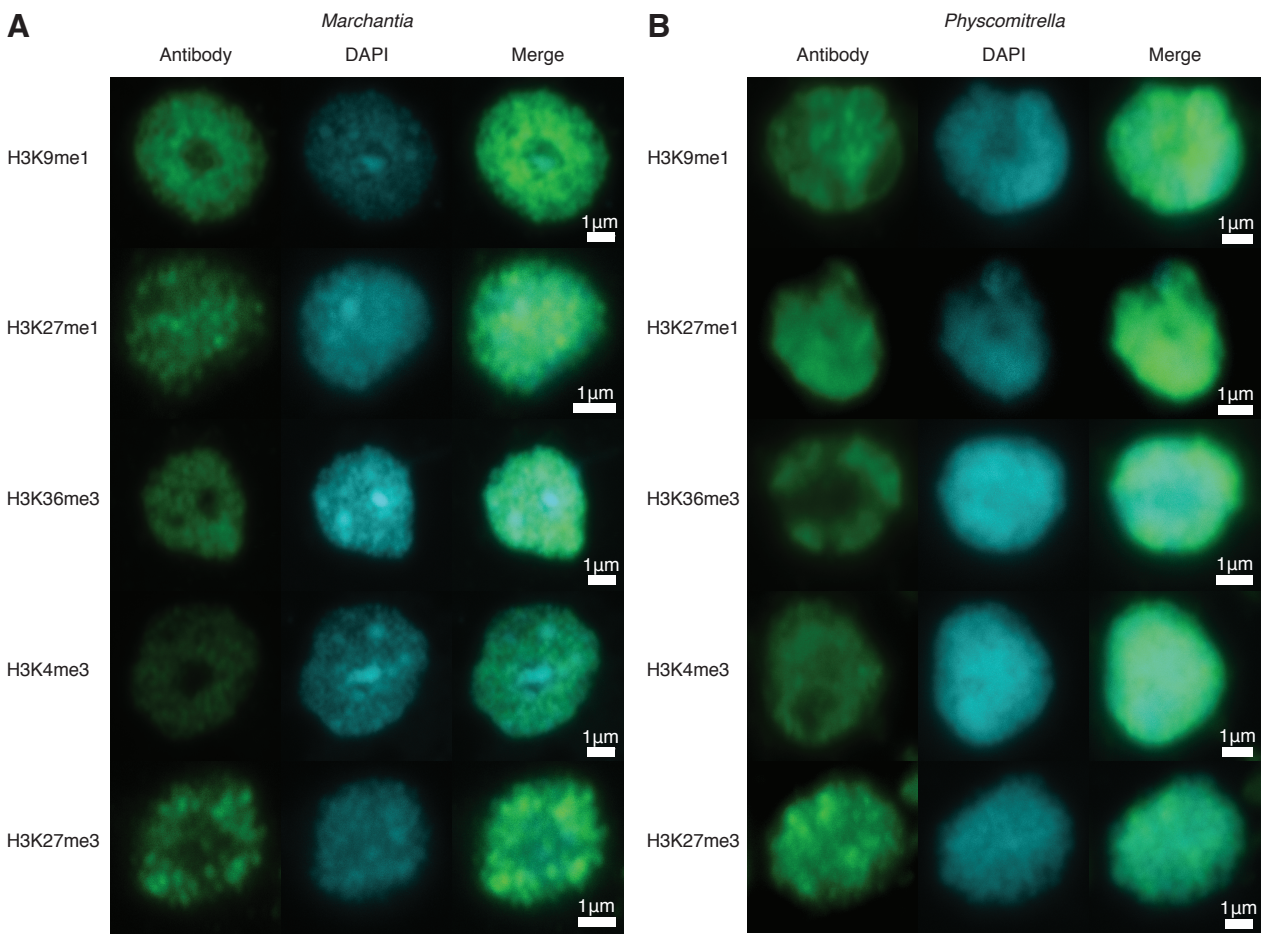

A
Marchantia
Antibody
DAPI
Merge
H3K9me1
H3K27me1
H3K36me3
H3K4me3
H3K27me3
1μm
B
Physcomitrella
Antibody
DAPI
Merge
H3K9me1
H3K27me1
H3K36me3
H3K4me3
H3K27me3
1μm

Figure S4. *Marchantia* and *Physcomitrella* nuclei immunostaining. Related to Figure 1.

(A) Immunostaining of isolated *Marchantia* nuclei. Green is the indicated chromatin mark. Blue is DAPI-stained DNA.

(B) Immunostaining of isolated *Physcomitrella patens* nuclei. Green is the indicated chromatin mark. Blue is DAPI-stained DNA.

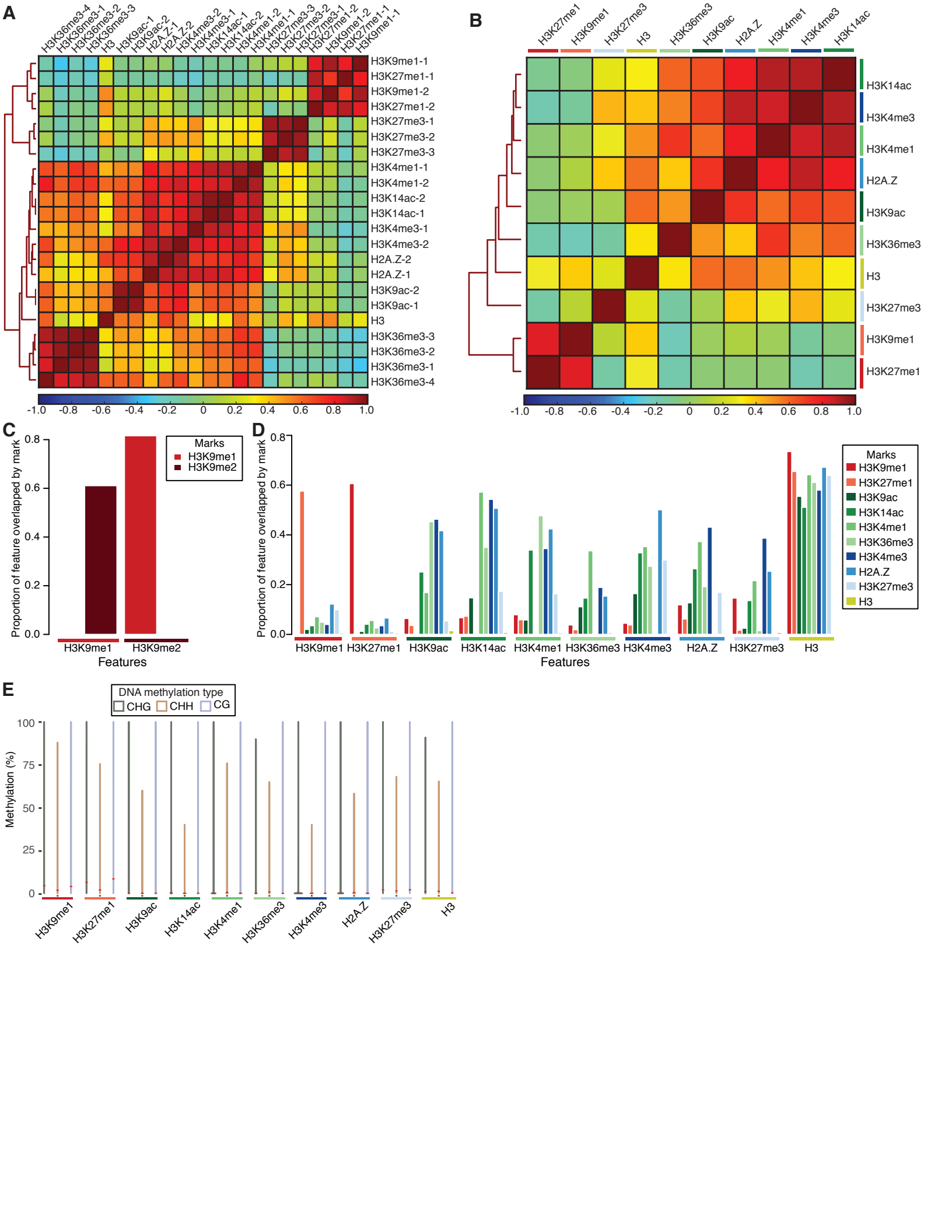

A
H3K36me3-4
H3K36me3-1
H3K36me3-2
H3K36me3-3
H3
H3K9ac-1
H3K9ac-2
H2A.Z-1
H2A.Z-2
H3K4me3-2
H3K4me3-1
H3K14ac-1
H3K14ac-2
H3K4me1-2
H3K4me1-1
H3K27me3-3
H3K27me3-2
H3K27me1-1
H3K9me1-2
H3K27me1-2
H3K9me1-1
H3K9me1-1
H3K27me1-1
H3K9me1-2
H3K27me1-2
H3K27me3-1
H3K27me3-2
H3K27me3-3
H3K4me1-1
H3K4me1-2
H3K14ac-2
H3K14ac-1
H3K4me3-1
H3K4me3-2
H2A.Z-2
H2A.Z-1
H3K9ac-2
H3K9ac-1
H3
H3K36me3-3
H3K36me3-2
H3K36me3-1
H3K36me3-4
-1.0 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1.0
B
H3K27me1
H3K9me1
H3K27me3
H3
H3K36me3
H3K9ac
H2A.Z
H3K4me1
H3K4me3
H3K14ac
H3K14ac
H3K4me3
H3K4me1
H2A.Z
H3K9ac
H3K36me3
H3
H3K27me3
H3K9me1
H3K27me1
-1.0 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1.0
C
Proportion of feature overlapped by mark
0.0 0.2 0.4 0.6 0.8
Marks
H3K9me1
H3K9me2
H3K9me1 H3K9me2
Features
D
Proportion of feature overlapped by mark
0.0 0.2 0.4 0.6 0.8
Marks
H3K9me1
H3K27me1
H3K9ac
H3K14ac
H3K4me1
H3K36me3
H3K4me3
H2A.Z
H3K27me3
H3
H3K9me1 H3K27me1 H3K9ac H3K14ac H3K4me1 H3K36me3 H3K4me3 H2A.Z H3K27me3 H3
Features
E
DNA methylation type
CHG CHH CG
Methylation (%)
0 25 50 75 100
H3K9me1 H3K27me1 H3K9ac H3K14ac H3K4me1 H3K36me3 H3K4me3 H2A.Z H3K27me3 H3

Figure S5. Distribution of chromatin marks in the *Marchantia* genome. Related to Figures 2, 3 and 4.

(A) Pearson correlation heatmap of CUT&RUN biological replicates. Colours represent the correlation coefficient, with red for high similarity and blue for low similarity. Hierarchical clustering shown to the left of the heatmap.

(B) Pearson correlation heatmap of merged CUT&RUN samples. Colours represent the correlation coefficient, with red for high similarity and blue for low similarity. Hierarchical clustering shown to the left of the heatmap.

(C) Proportion of H3K9me1 and H3K9me2 peaks overlapping each other in *Arabidopsis*. The total length of overlapping peaks was divided by the total length of feature peaks (x-axis) to determine the proportion of feature lengths overlapped by marks.

(D) Proportion of chromatin mark peaks overlapped by other chromatin mark peaks. The total length of overlapping chromatin mark peaks was divided by the total length of peaks of chromatin marks (features; along x axis) to determine the proportion of feature lengths overlapped by other chromatin marks.

(E) DNA methylation levels over chromatin mark peaks. Methylation percentage calculated per chromatin mark peak. Width relative to density of peaks. Red dots indicate median methylation values.

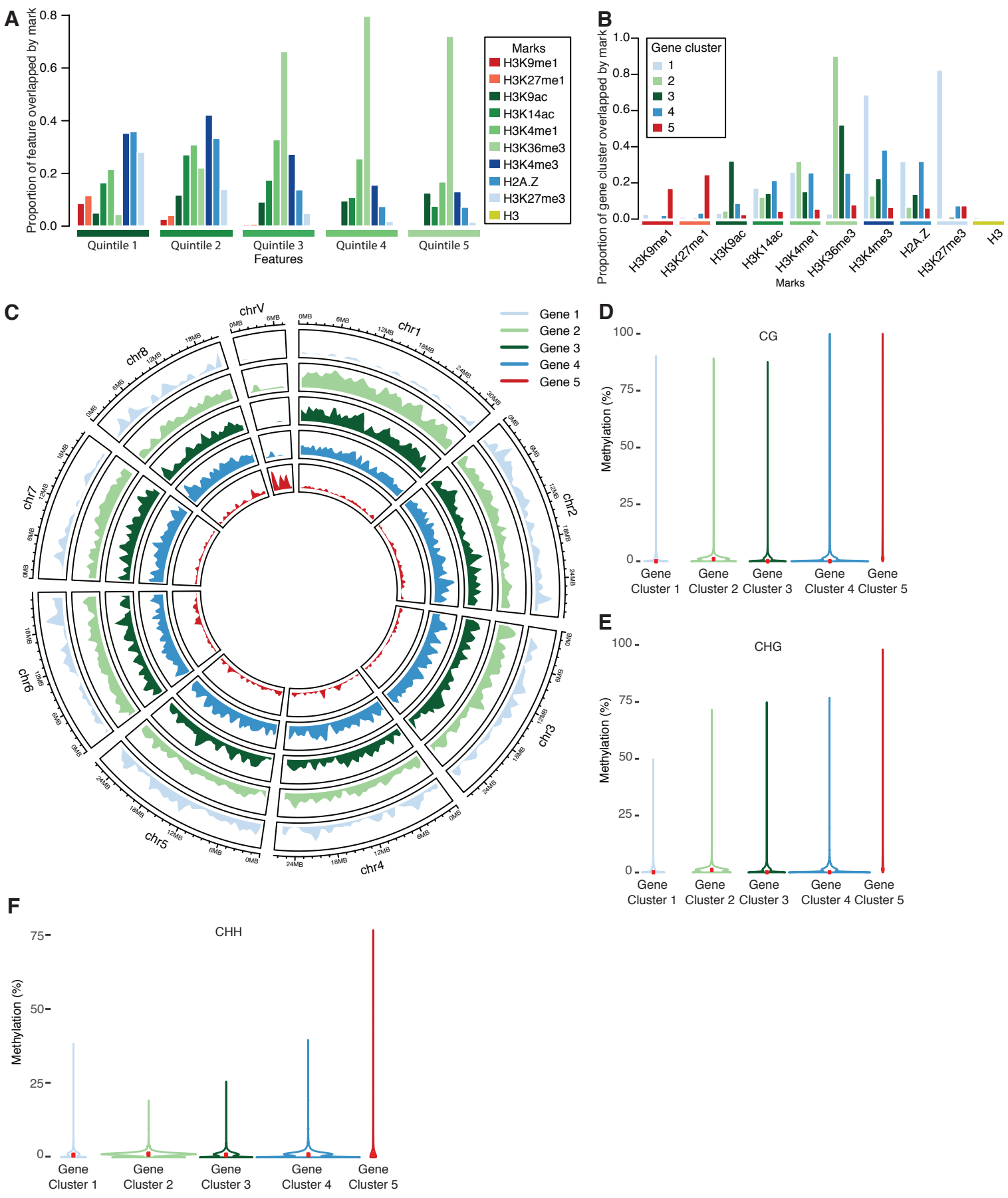

A
Proportion of feature overlapped by mark
0.0
0.2
0.4
0.6
0.8
Quintile 1
Quintile 2
Quintile 3
Quintile 4
Quintile 5
Features
Marks
H3K9me1
H3K27me1
H3K9ac
H3K14ac
H3K4me1
H3K36me3
H3K4me3
H2A.Z
H3K27me3
H3
B
Proportion of gene cluster overlapped by mark
0.0
0.2
0.4
0.6
0.8
1.0
Gene cluster
1
2
3
4
5
H3K9me1
H3K27me1
H3K9ac
H3K14ac
H3K4me1
H3K36me3
H3K4me3
H2A.Z
H3K27me3
H3
Marks
C
chrV
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
Gene 1
Gene 2
Gene 3
Gene 4
Gene 5
D
CG
Methylation (%)
0
25
50
75
100
Gene Cluster 1
Gene Cluster 2
Gene Cluster 3
Gene Cluster 4
Gene Cluster 5
E
CHG
Methylation (%)
0
25
50
75
100
Gene Cluster 1
Gene Cluster 2
Gene Cluster 3
Gene Cluster 4
Gene Cluster 5
F
CHH
Methylation (%)
0
25
50
75
Gene Cluster 1
Gene Cluster 2
Gene Cluster 3
Gene Cluster 4
Gene Cluster 5

Figure S6. Association of chromatin marks with genes. Related to Figure 3.

(A) Proportion of genes per expression quintile overlapped by chromatin mark peaks. The total length of chromatin mark peaks overlapping genes was divided by the total length of genes per quintile to determine each proportion. Quintiles correspond to transcript per million values as follows: 1: 0-0.073; 2: 0.073-2.013; 3: 2.013-12.410; 4: 12.410-33.950; 5: 33.950-23567.63.

(B) Proportion of gene clusters overlapped by chromatin mark peaks. The total length of chromatin mark peaks overlapping genes per gene cluster was divided by the total length of genes per gene cluster to determine each proportion.

(C) Circos plot of gene cluster distribution across the genome. Each band shows the density of genes per gene cluster per chromosome, relative to the greatest density per band.

(D) DNA CG methylation levels of gene clusters. Methylation percentage calculated per gene in each gene cluster. Width relative to density of genes. Red dots indicate median methylation values.

(E) DNA CHG methylation levels of gene clusters. Methylation percentage calculated per gene in each gene cluster. Width relative to density of genes. Red dots indicate median methylation values.

(F) DNA CHH methylation levels of gene clusters. Methylation percentage calculated per gene in each gene cluster. Width relative to density of genes. Red dots indicate median methylation values.

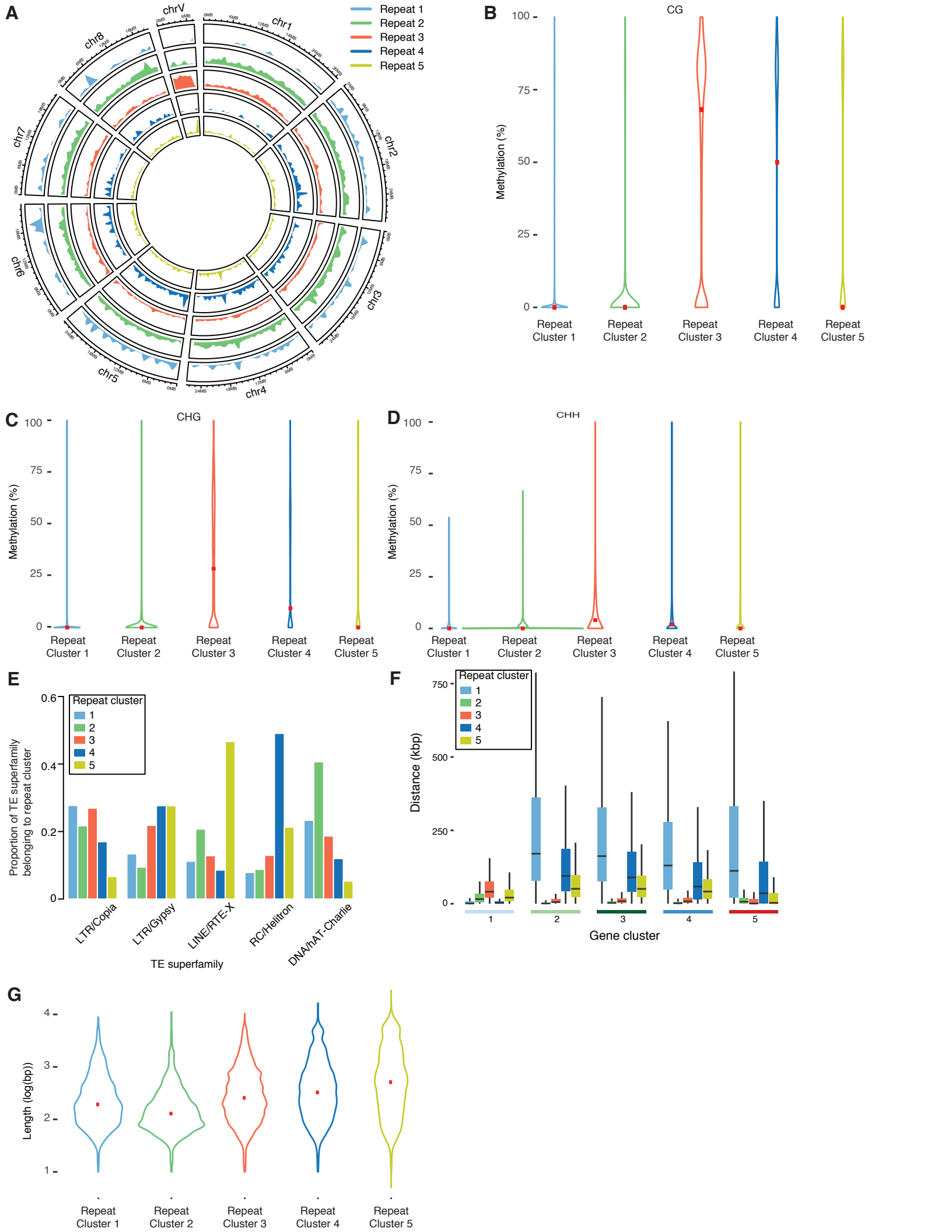

A
chrV
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
Repeat 1
Repeat 2
Repeat 3
Repeat 4
Repeat 5
B
CG
Methylation (%)
100
75
50
25
0
Repeat Cluster 1
Repeat Cluster 2
Repeat Cluster 3
Repeat Cluster 4
Repeat Cluster 5
C
CHG
Methylation (%)
D
CHH
Methylation (%)
E
Repeat cluster
1
2
3
4
5
Proportion of TE superfamily belonging to repeat cluster
0.6
0.4
0.2
0
LTR/Copia
LTR/Gypsy
LINE/RTE-X
RC/Helitron
DNA/hAT-Charlie
TE superfamily
F
Repeat cluster
Distance (kbp)
750
500
250
0
1
2
3
4
5
Gene cluster
G
Length (log(bp))
4
3
2
1
Repeat Cluster 1
Repeat Cluster 2
Repeat Cluster 3
Repeat Cluster 4
Repeat Cluster 5

Figure S7. Association of chromatin marks with transposons. Related to Figure 4.

(A) Circos plot of repeat cluster distribution across the genome. Each band shows the density of transposons per repeat cluster per chromosome, relative to the greatest density per band.

(B) DNA CG methylation levels of repeat clusters. Methylation percentage calculated per transposon in each repeat cluster. Width relative to density of transposons. Red dots indicate median methylation values.

(C) DNA CHG methylation levels of repeat clusters. Methylation percentage calculated per transposon in each repeat cluster. Width relative to density of transposons. Red dots indicate median methylation values.

(D) DNA CHH methylation levels of repeat clusters. Methylation percentage calculated per transposon in each repeat cluster. Width relative to density of transposons. Red dots indicate median methylation values.

(E) Proportion of transposon superfamily length belonging to repeat clusters. The total length of transposon superfamilies belonging to repeat clusters was divided by the total length each repeat cluster and scaled per transposon superfamily to determine each proportion.

(F) Boxplot of distances between each gene and the nearest transposon per repeat cluster. Briefly each gene is compared to all transposons belonging to a repeat cluster to find its nearest neighbor. Genes are divided based on the gene cluster they belong to. Distances in kilobases (kbp). Coloured boxes represent interquartile range and lines represent median values. Outliers not shown.

(G) Length of repeats per cluster in log scale. Width relative to density of genes. Red dots indicate median lengths of repeats per cluster.

Sample	Total reads sequenced	Total mapped reads	Total mapped reads after removing PCR duplicates	Total Hi-C reads after filtering
Tak-1 Hi-C rep1	111916105	70619675	60276489	48658854
Tak-1 Hi-C rep2	133115486	81254888	50779358	41369547

Table S2. Statistics related to Hi-C data. Related to STAR Methods.

Supplemental References

[S1] Robinson, J.T., Turner, D., Durand, N.C., Thorvaldsdóttir, H., Mesirov, J.P., and Aiden, E.L. (2018). Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. Cell Systems *6*, 256-258 e1.

[S2] Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. Cell *171*, 287-304 e215.

[S3] Yamato, K.T., Ishizaki, K., Fujisawa, M., Okada, S., Nakayama, S., Fujishita, M., Bando, H., Yodoya, K., Hayashi, K., Bando, T., et al. (2007). Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc Natl Acad Sci U S A *104*, 6472-6477.

[S4] Fujisawa, M., Nakayama, S., Nishio, T., Fujishita, M., Hayashi, K., Ishizaki, K., Kajikawa, M., Yamato, K.T., Fukuzawa, H., and Ohyama, K. (2003). Evolution of ribosomal DNA unit on the X chromosome independent of autosomal units in the liverwort Marchantia polymorpha. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology *11*, 695-703.