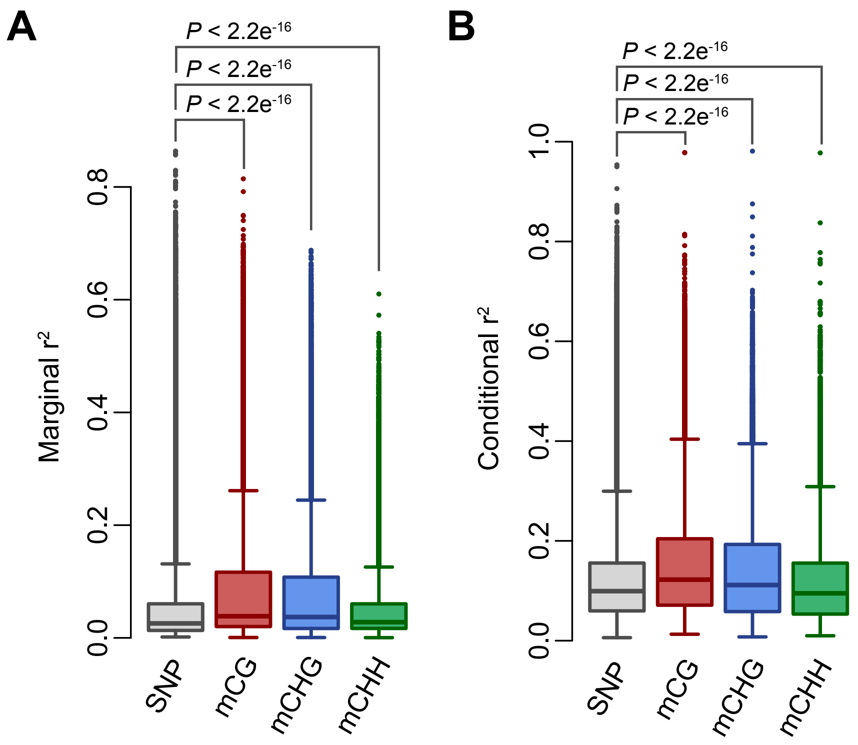
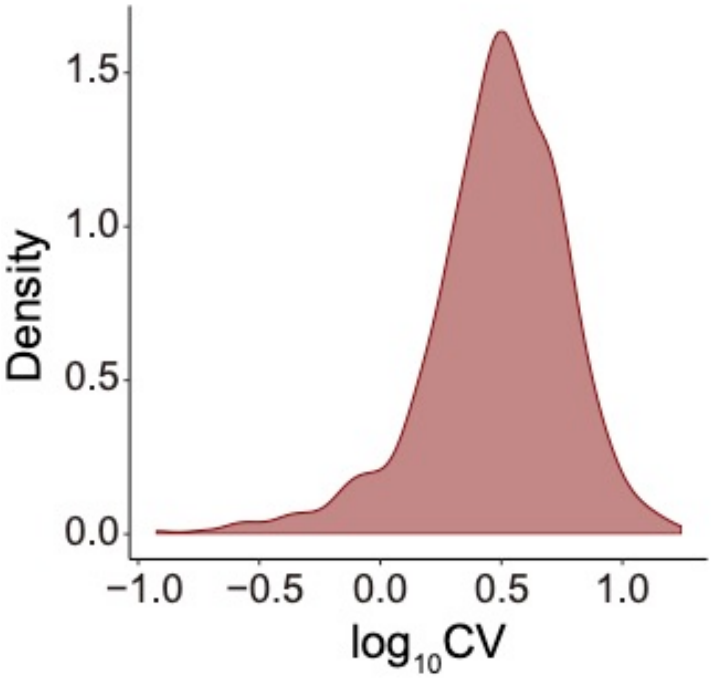


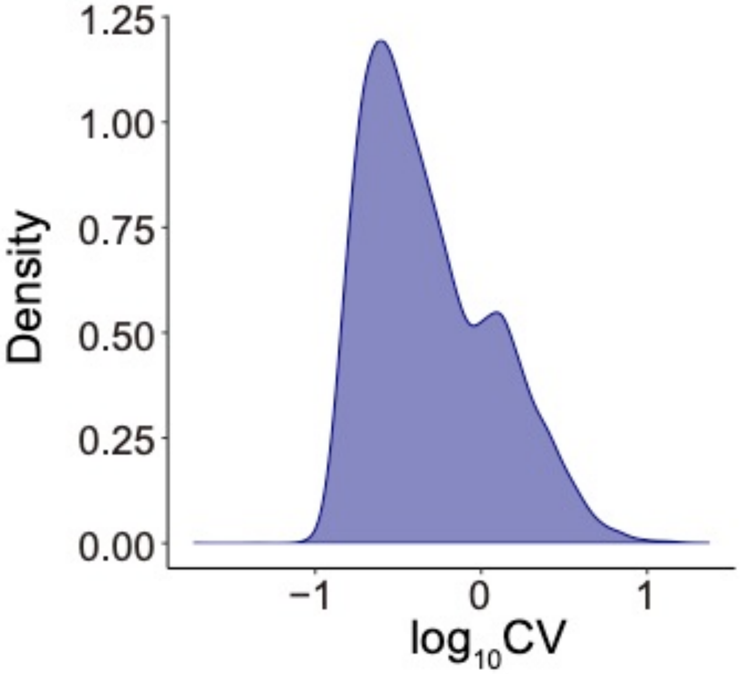
Supplemental Fig. S1.

The distribution of r^2 calculated from LMM in genes related to SMPs (or SNPs). (*A*) The distribution of marginal r^2. (*B*) The distribution of conditional r^2. In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The significant differences in r^2 among genes associated with SNPs and the SMP groups were evaluated using the Wilcoxon rank sum test.



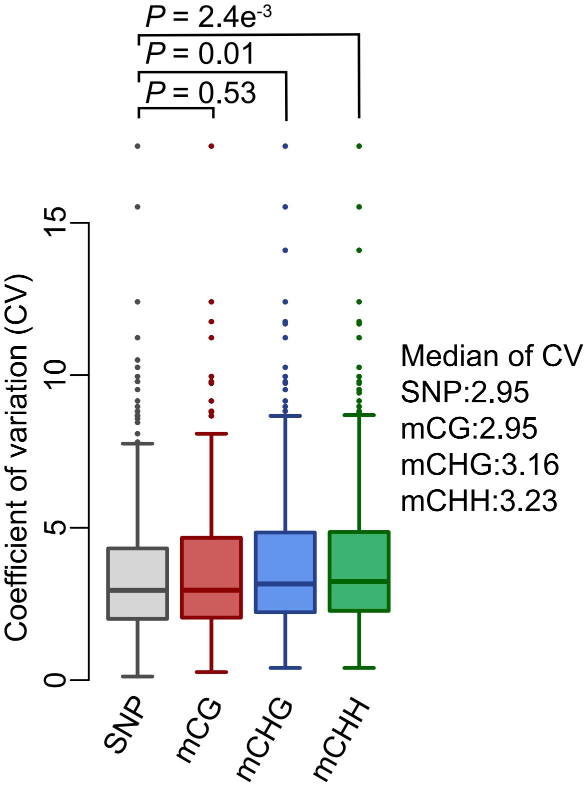
Supplemental Fig. S2.

Coefficients of variation (CVs) for 669 TE genes from 620 accessions.



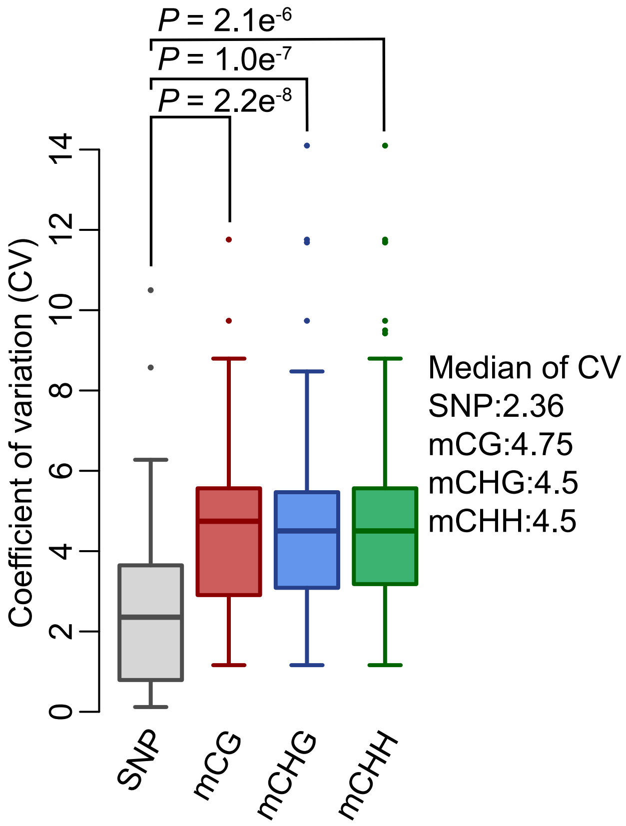
Supplemental Fig. S3.

Coefficients of variation (CVs) for 23,361 the other genes (excluding TE genes) from 620 accessions.



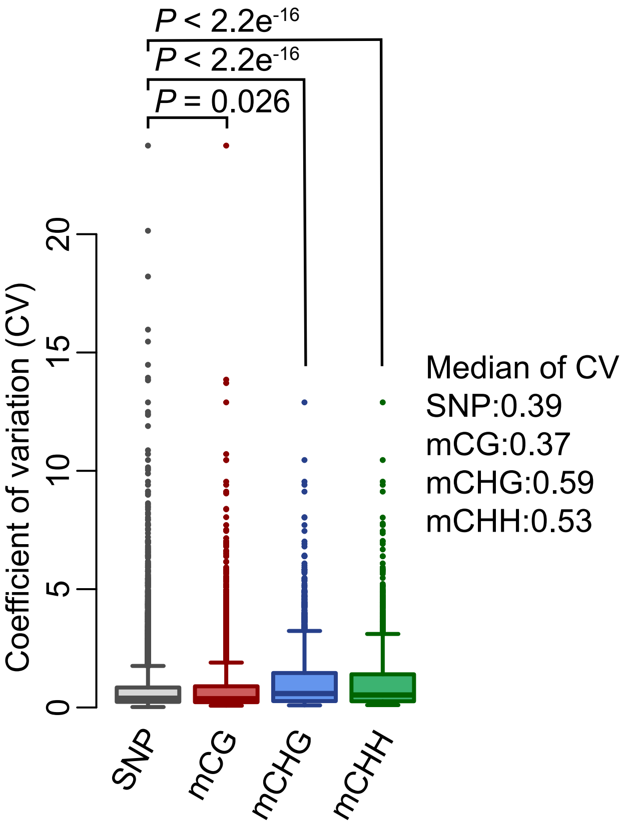
Supplemental Fig. S4.

CVs of TE genes associated with SNPs, mCGs, mCHGs, and mCHHs. In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The significant differences in CVs among genes associated with SNPs and the SMP groups were evaluated using the Wilcoxon rank sum test.



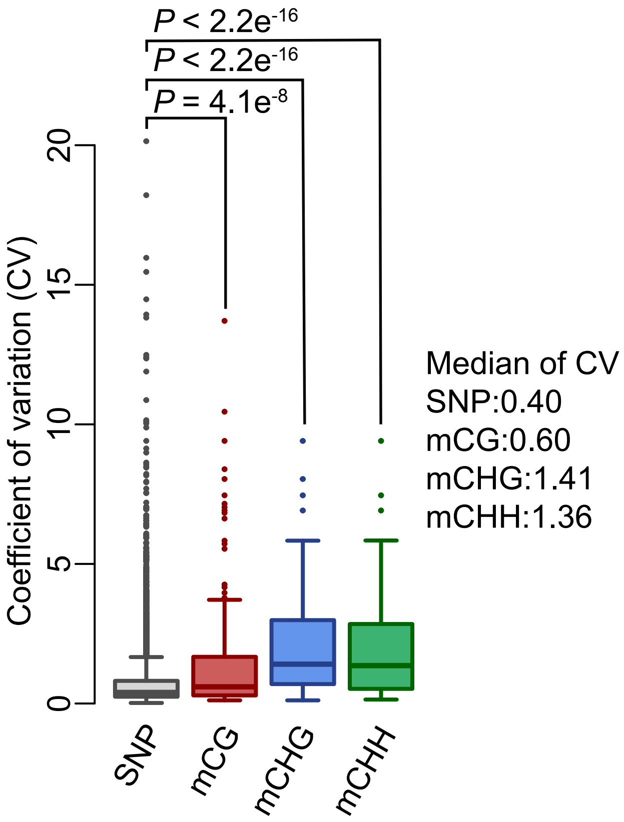
Supplemental Fig. S5.

CVs of TE genes associated with only SNPs, only mCGs, only mCHGs, and only mCHHs. In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The significant differences in CVs among genes associated with SNPs and the SMP groups were evaluated using the Wilcoxon rank sum test.



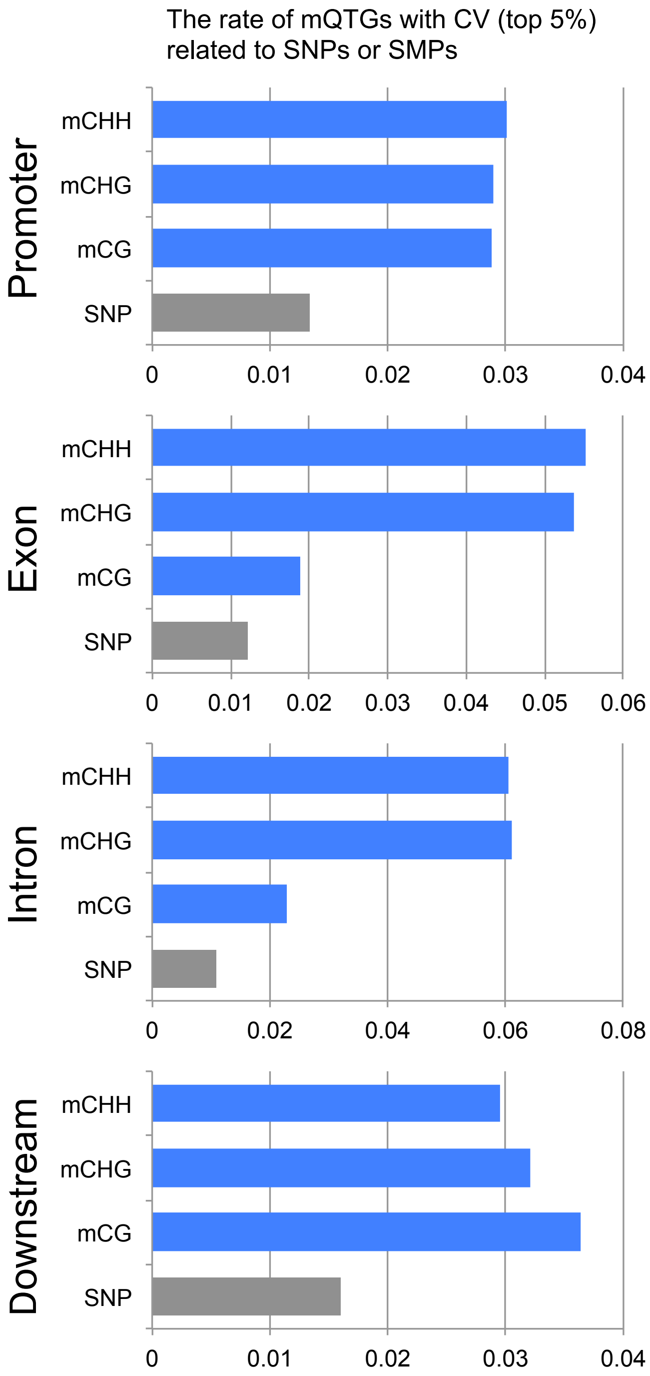
Supplemental Fig. S6.

CVs of the other genes (excluding TE genes) associated with SNPs, mCGs, mCHGs, and mCHHs. In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The significant differences in CVs among genes associated with SNPs and the SMP groups were evaluated using the Wilcoxon rank sum test.



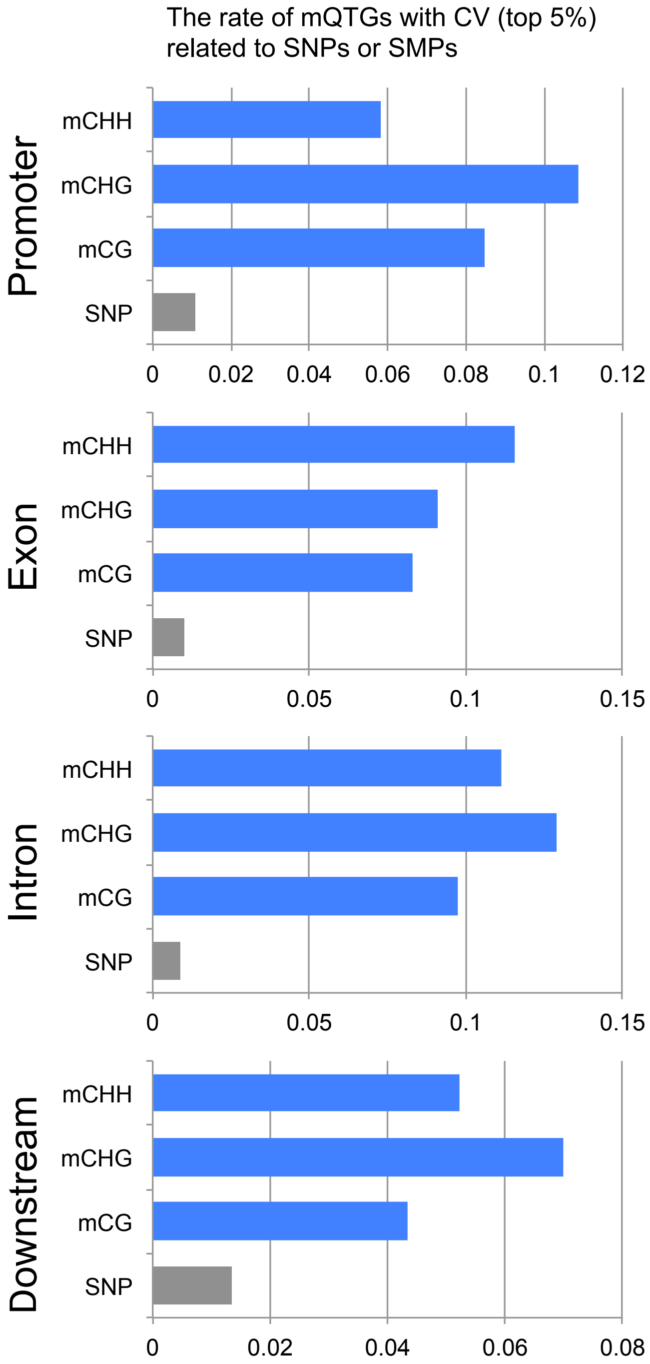
Supplemental Fig. S7.

CVs of the other genes (excluding TE genes) associated with only SNPs, only mCGs, only mCHGs, and only mCHHs. In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The significant differences in CVs among genes associated with SNPs and the SMP groups were evaluated using the Wilcoxon rank sum test.



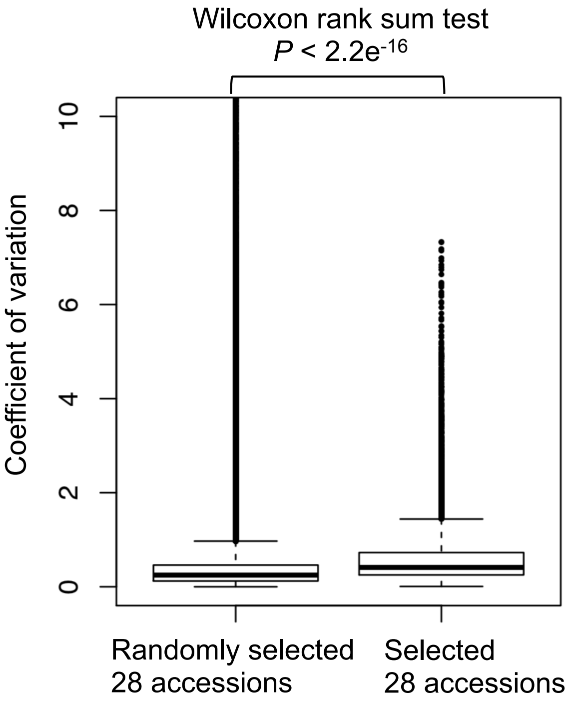
Supplemental Fig. S8.

The rate of mQTGs with CV (top 5%) related to SNPs or SMPs at 620 accessions. This rate is the number of mQTGs with CV (top 5%) associated with SNPs (or SMPs)/the number of genes associated with SNPs (or SMPs).



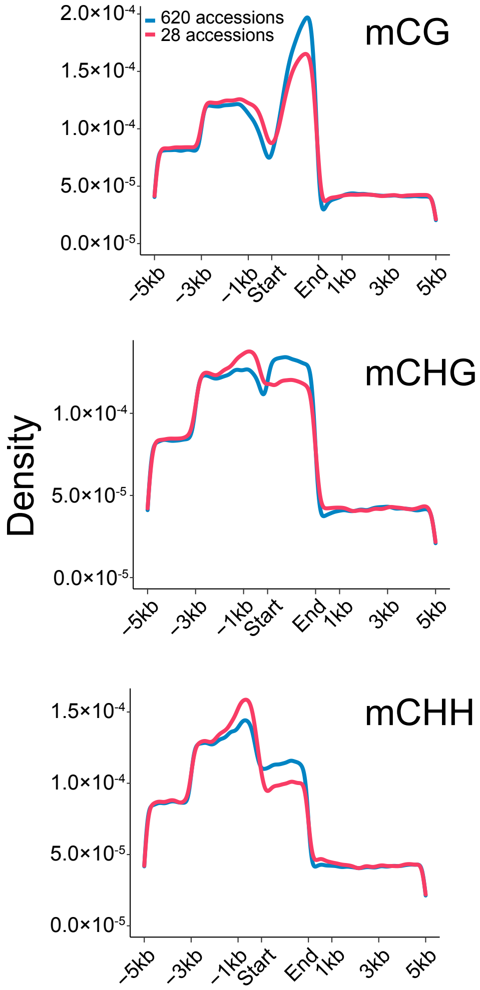
Supplemental Fig. S9.

The rate of mQTGs with CV (top 5%) related to only SNPs or SMPs at 620 accessions. This rate is the number of mQTGs with CV (top 5%) associated with only SNPs (or SMPs)/the number of genes associated with only SNPs (or SMPs).



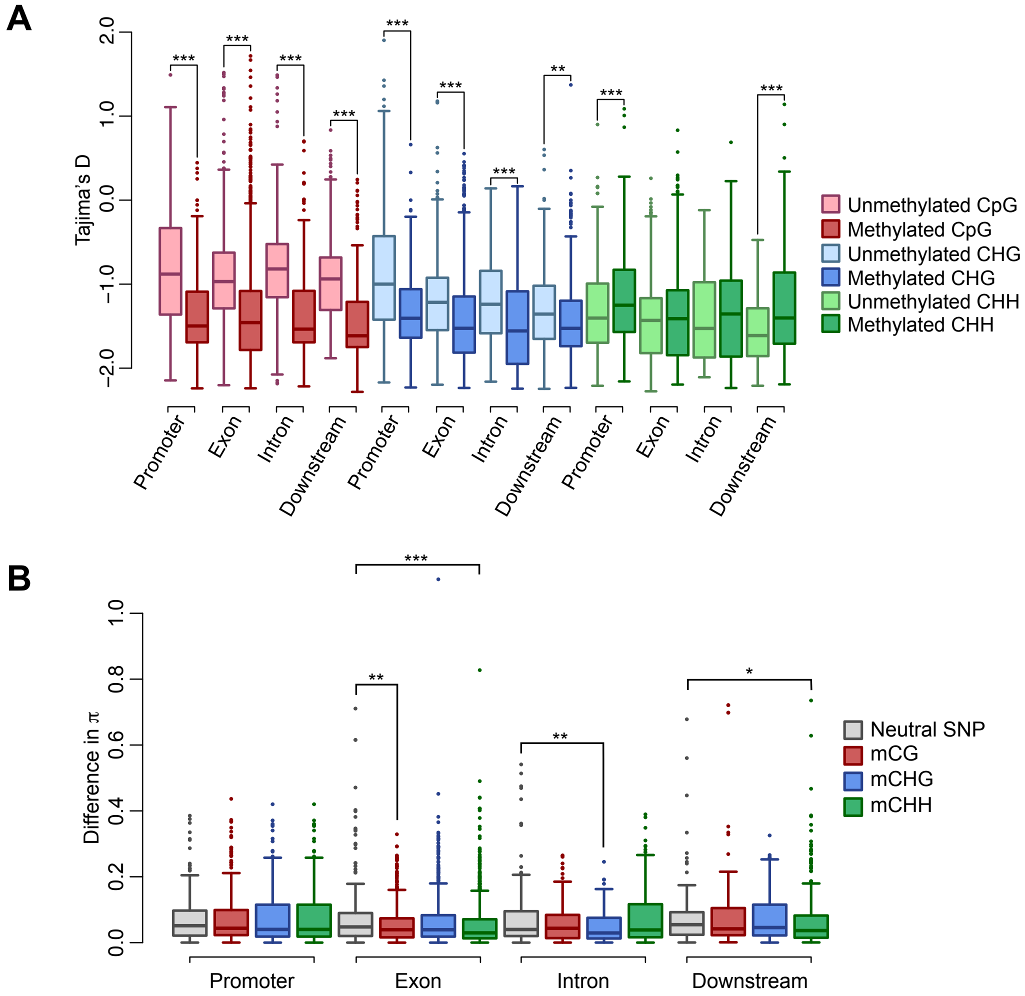
Supplemental Fig. S10.

Comparison of coefficient of variation (CV) between randomly selected 28 accessions and manually selected 28 accessions. We selected 28 accessions from the 75 accessions. The CVs of gene expression were calculated at 21,957 genes among the 28 accessions. We repeated these steps at 10,000 times. Randomly selected 28 accessions represents the all of the CVs.



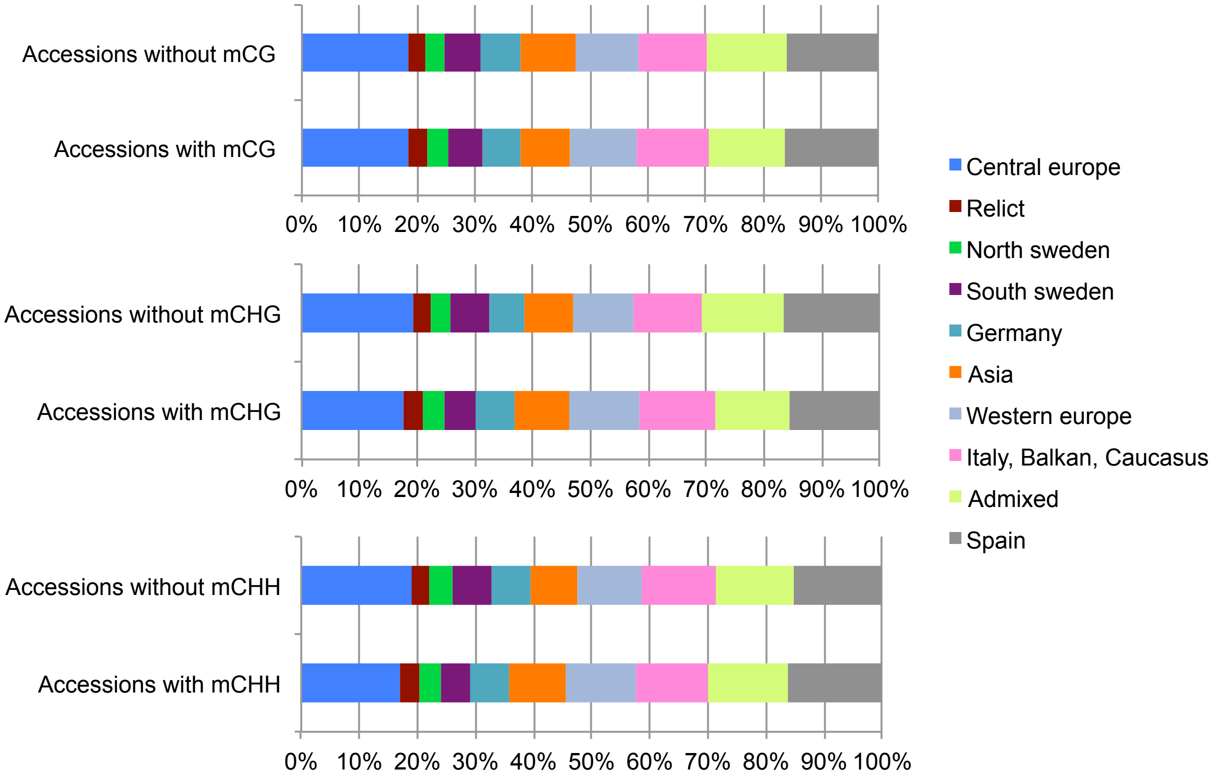
Supplemental Fig. S11.

The distribution of each type of SMPs around all of protein coding genes. The blue line represents distribution of SMPs in the global 620 accessions. The red line represents distribution of SMPs in the characteristic 28 accessions. The X axis represents the distance between location of SMPs and start site (or end site) of a gene.



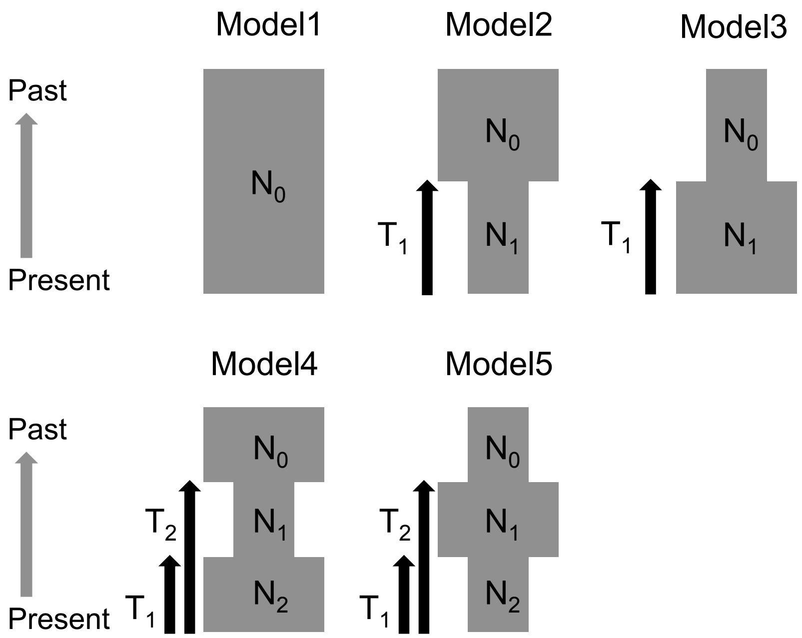
Supplemental Fig. S12.

The differences in nucleotide diversity around SMPs associated with expression of mQTGs (top 5% CV). (*A*) Tajima's *D* around SMPs associated with expression of mQTGs with high CV (top 5%). (*B*) The differences in nucleotide diversity (π) around SMPs associated with expression of mQTGs with high CV (top 5%). The differences between the two group were tested by Wilcoxon rank sum test at each location; ****P* < 0.001, ***P* < 0.01, **P* < 0.05.



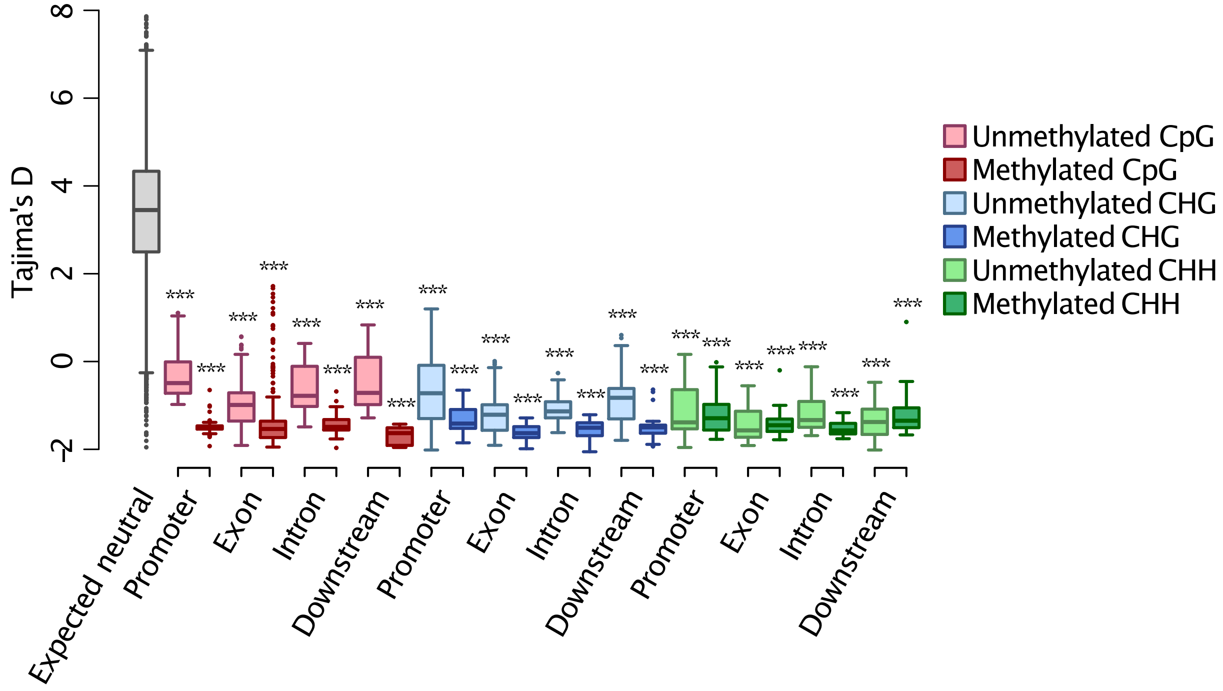
Supplemental Fig. S13.

The proportion of accessions classified by geographic distributions in accessions with and without SMPs associated with expression of mQTGs (top 1% CV). The 620 accessions were classified into ten groups (Central europe, Relict, North sweden, South sweden, Germany, Asia, Western europe, (Italy, Balkan, Caucasus), Admixed, Spain). The classification of accessions is based on 1001 Genomes Consortium. At each of the 732 SMPs (330 mCGs, 214 mCHGs, and 188 mCHHs) associated with mQTGs having the 1% highest CVs, we counted the number of accessions with and without SMPs. The graphs show the proportion of the accessions classified by geographic distributions in the total number of the accessions with or without SMPs.



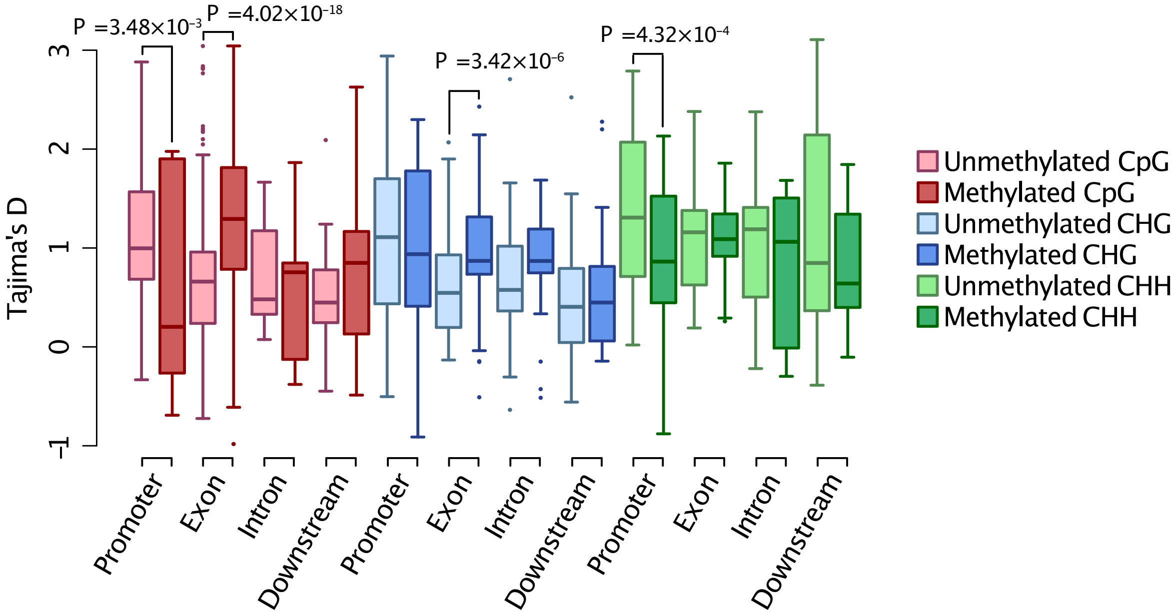
Supplemental Fig. S14.

The demographic models for Fastsimcoal26. The N_0, N_1 and N_2 represent effective population size. The T_1 and T_2 represent the timing of events when population size increased or decreased.



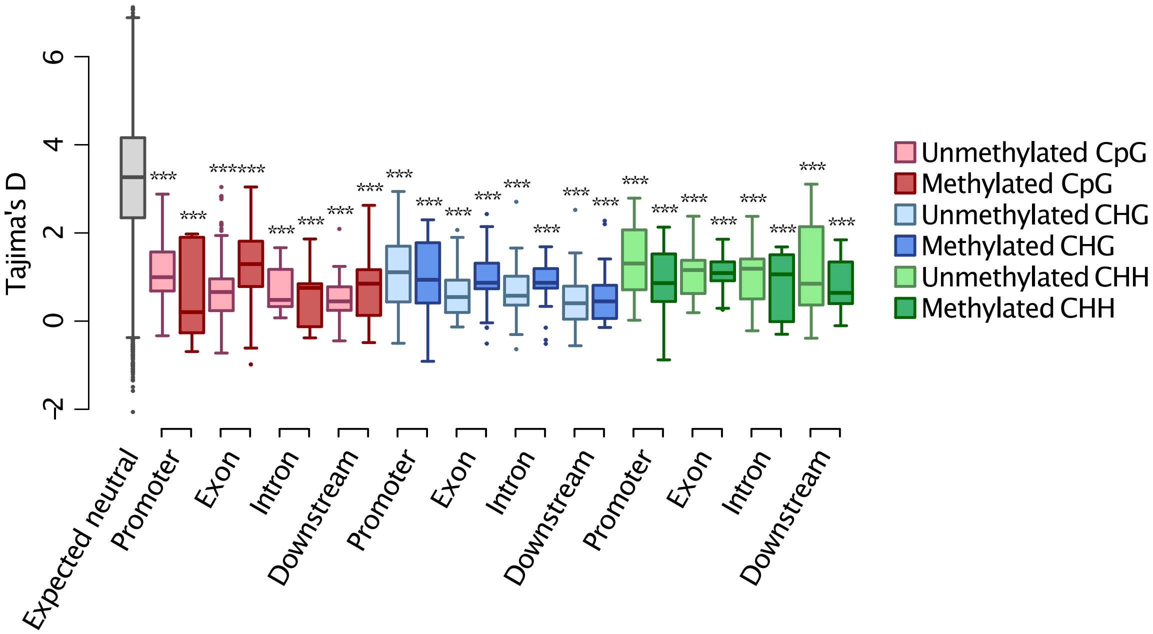
Supplemental Fig. S15.

Tajima's *D* values near SMPs associated with mQTG expression compared with expected neutral Tajima's *D* values. The *X*-axis represents SMP locations, and the *Y*-axis represents Tajima's *D* values. The boxes represent Tajima's *D* values of accessions with no DNA methylation or accessions with DNA methylation at SMP sites associated with mQTGs having high CVs (top 1%). In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The differences between expected neutral and the other groups were determined using a Wilcoxon rank sum test; ***FDR < 0.001, **FDR < 0.01, *FDR < 0.05.



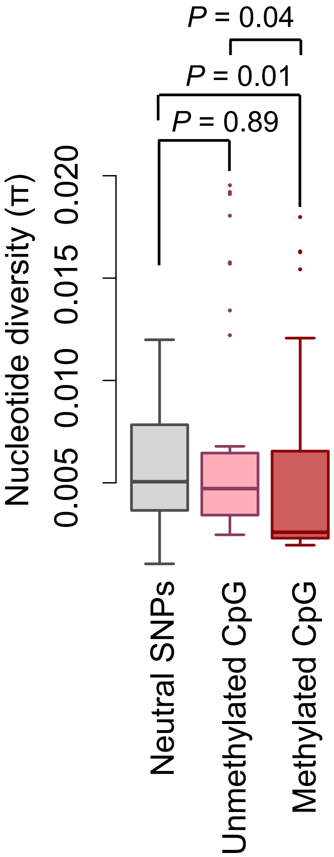
Supplemental Fig. S16.

Tajima's *D* values near SMPs associated with mQTG expression calculated in 100 SNPs windows. The *X*-axis represents SMP locations, and the *Y*-axis represents Tajima's *D* values in 100 SNPs windows. The boxes represent Tajima's *D* values of accessions with no DNA methylation or accessions with DNA methylation at SMP sites associated with mQTGs having high CVs (top 1%). In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The differences between the two groups were determined at each location using a Wilcoxon rank sum test.



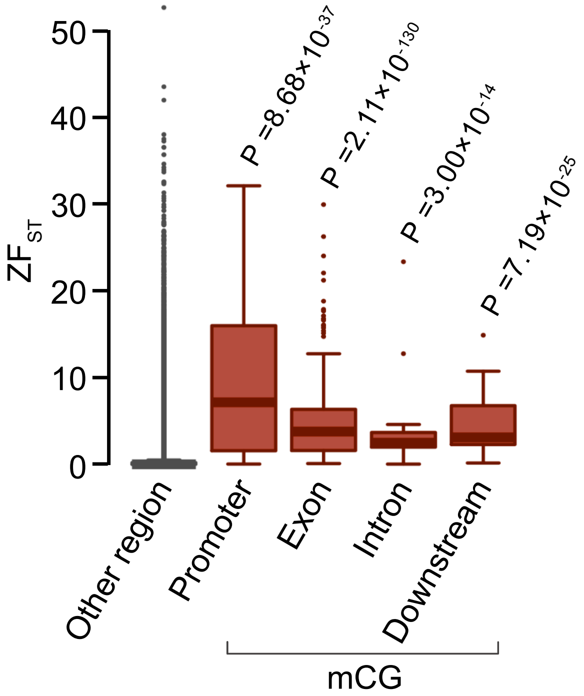
Supplemental Fig. S17.

Tajima’s *D* values near SMPs associated with mQTG expression compared with expected neutral Tajima’s *D* in 100 SNPs windows. The *X*-axis represents SMP locations, and the *Y*-axis represents Tajima’s *D* values in 100 SNPs windows. The boxes represent Tajima’s *D* values of accessions with no DNA methylation or accessions with DNA methylation at SMP sites associated with mQTGs having high CVs (top 1%). In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The differences between expected neutral and the other groups were determined using a Wilcoxon rank sum test; ***FDR < 0.001, **FDR < 0.01, *FDR < 0.05.



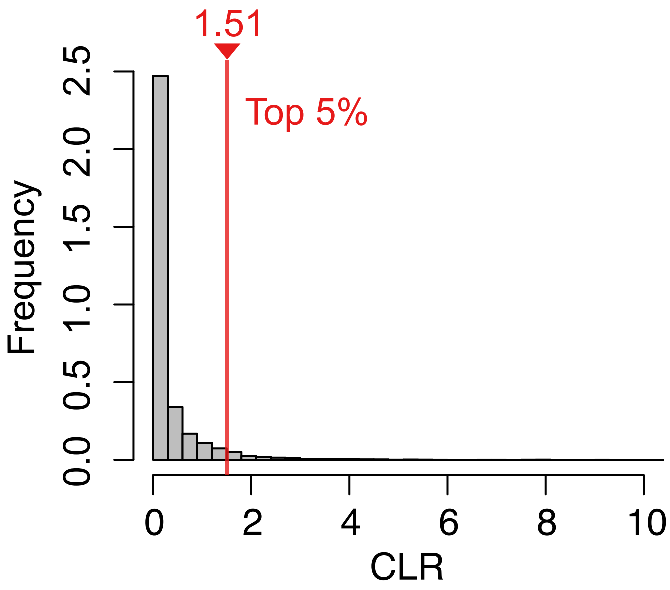
Supplemental Fig. S18.

Nucleotide diversity (π) around mCG associated with expression of mQTGs (top 1% CV) in promoter region. The gray bar shows the π values around neutral SNPs in promoter region. The pink bar shows the π values around accessions without mCG associated with expression of mQTGs (top 1% CV) in promoter region. The red bar shows the π values around accessions with mCG associated with expression of mQTGs (top 1% CV) in promoter region. The differences between the two groups were tested by Wilcoxon rank sum test.



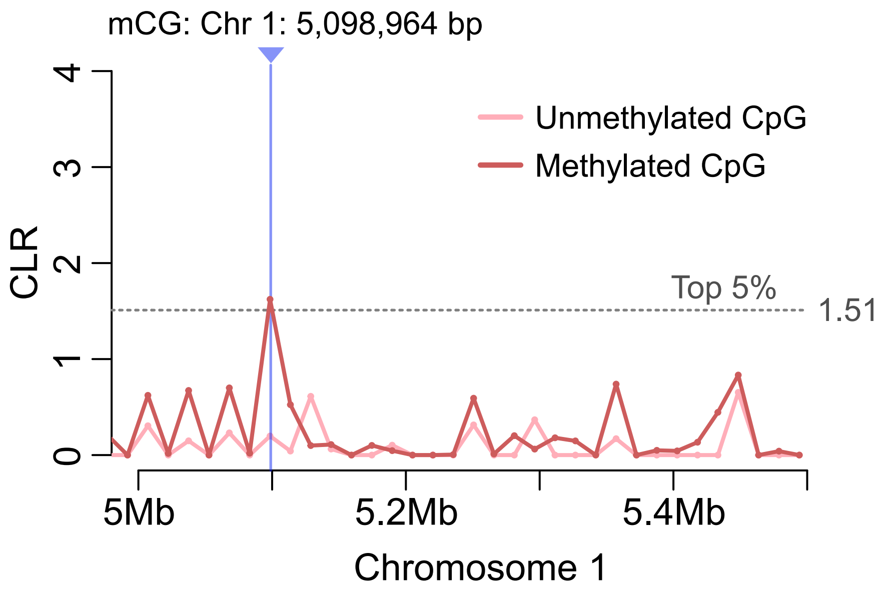
Supplemental Fig. S19.

Z-transformed F_{ST} (ZF_{ST}) near mCGs associated with mQTG expression with 10 kbp windows. The *X*-axis represents locations of the windows, and the *Y*-axis represents ZF_{ST} values. The ZF_{ST} values were calculated between accessions with mCGs associated with mQTGs having high CVs (top 1%) and accessions lacking the mCGs. The red boxes represent ZF_{ST} values of windows near the mCGs. In each box plot, the box represents the 25%–75% range, the middle line represents the median, the dotted line represents the 1%–99% range, and the outer circles represent outliers. The differences between windows near the mCGs and windows located on the other regions were determined using a Wilcoxon rank sum test.



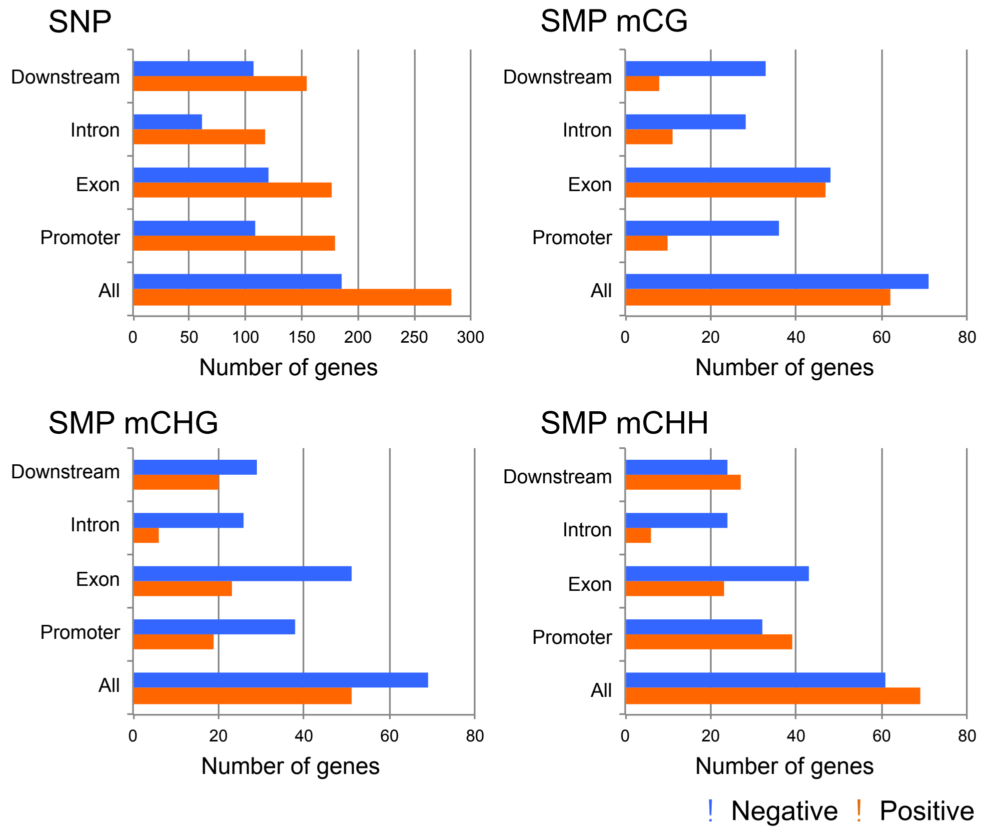
Supplemental Fig. S20.

The distribution of composite likelihood ratio (CLR) calculated by SweeD. The *X*-axis represents CLR calculated at any genome position. The red line represents the threshold of top 5% of whole genome CLR.



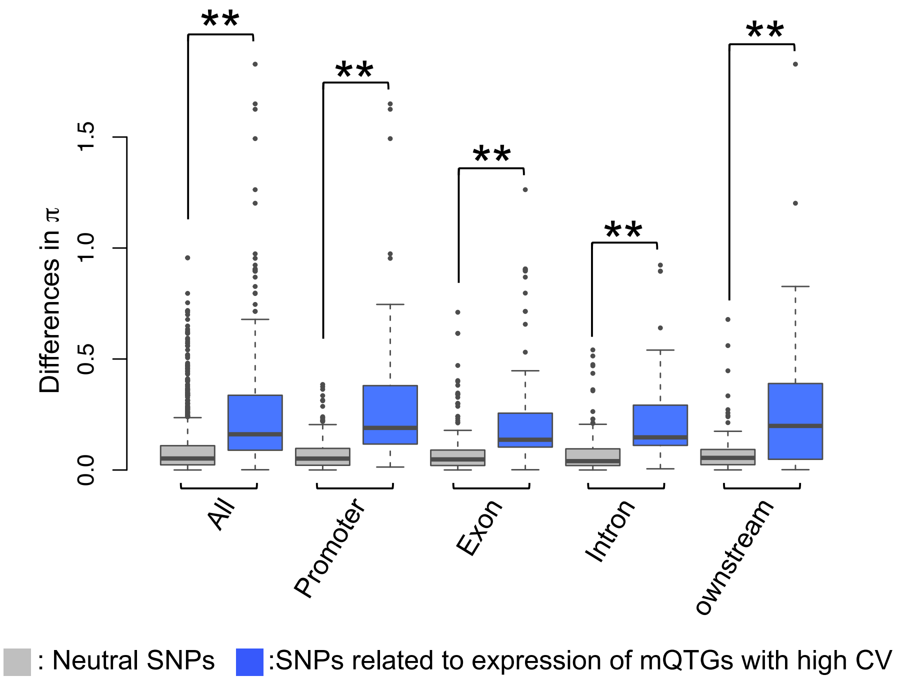
Supplemental Fig. S21.

Detection of selective sweep near the mCG located on Chr 1: 5,098,964. The *X*-axis represents locations of detection windows, and the *Y*-axis represents CLR calculated by SweeD. The red line represents CLR calculated from accessions with mCG at Chr 1: 5,098,964. The pink line represents CLR calculated from accessions lacking the mCG at Chr 1: 5,098,964. The dotted line represents the threshold of top 5% of whole genome CLR. The blue line represents position of Chr 1: 5,098,964.



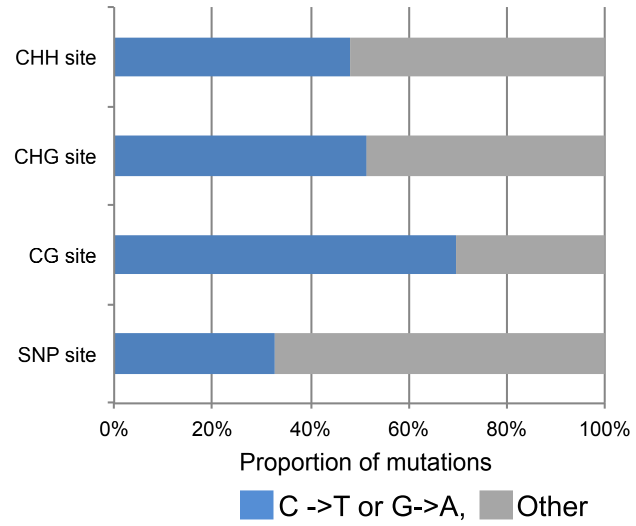
Supplemental Fig. S22.

Direction of association between SNPs (or SMPs) and expression of mQTGs with high CVs (top 5%). Blue bars represent the number of genes with negative association with SNPs (or SMPs). Orange bars represent the number of genes with positive association with SNPs (or SMPs).



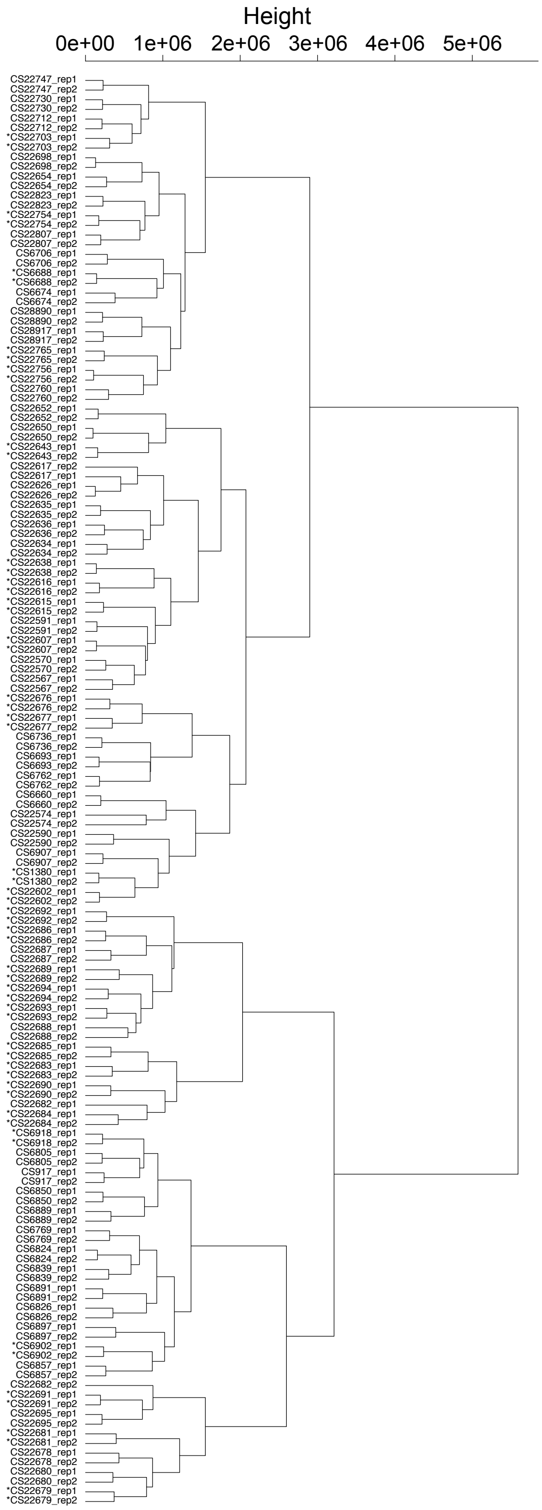
Supplemental Fig. S23.

Selective sweep for SNPs related to expression of mQTGs. The *X*-axis represents the locations of SNPs, whereas the *Y*-axis represents the differences in nucleotide diversity (π). In each box plot, the box represents the 25–75% range, the middle line represents the median, the dotted line represents the 1–99% range, and the outer circles are outliers. Gray boxes represent the π differences of neutral SNPs. Blue boxes represent the π differences of SNPs associated with mQTGs with high CV (top 1%). The differences between the neutral SNPs and the SNPs associated with mQTGs were tested by Wilcoxon rank sum test at each locations; $^{**}P < 0.01$, $^{*}P < 0.05$.



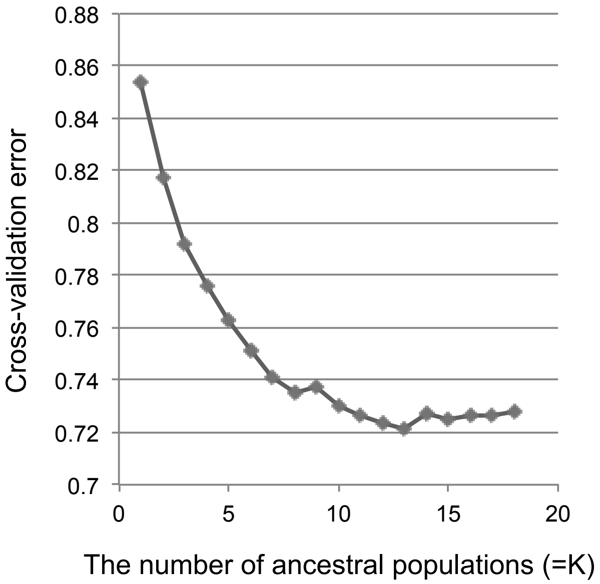
Supplemental Fig. S24.

Direction of mutations at methylated sites. Blue bars represent the rate of mutations C -> T or G ->A. Gray bars represent the rate of the other mutations.



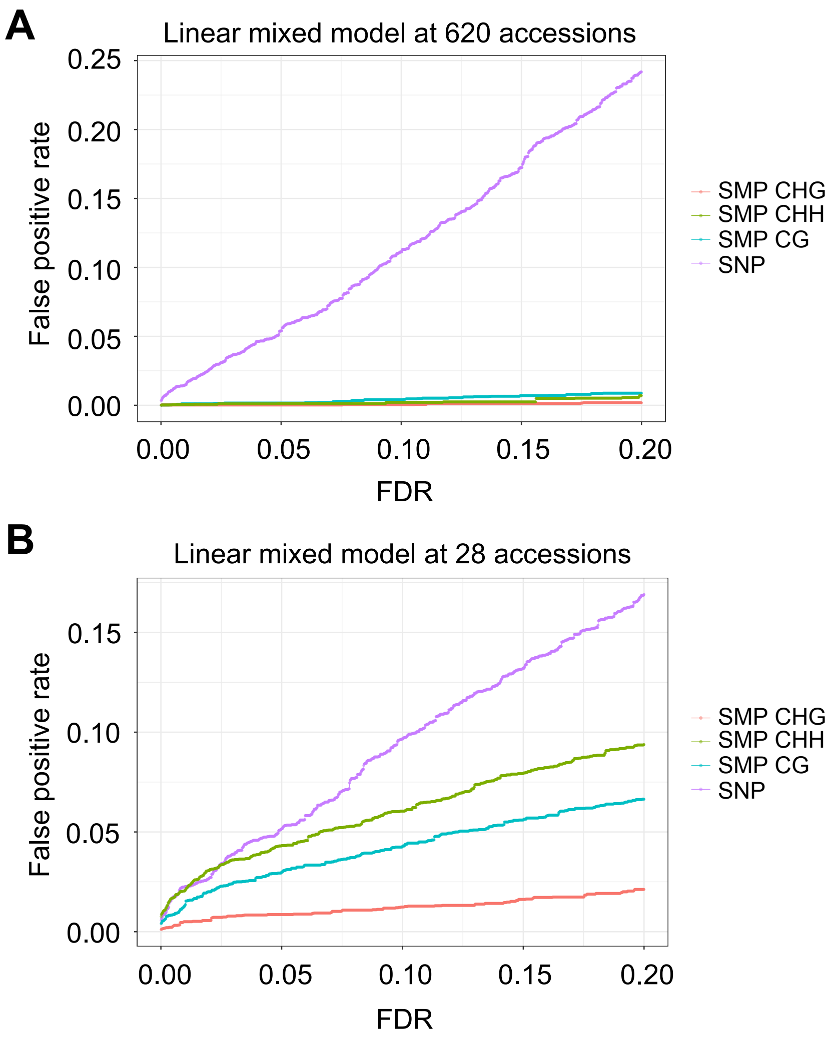
Supplemental Fig. S25.

Clustering of 75 accessions based on expression of 21,957 genes. The Clustering dendrogram was constructed by Word’s method. Asterisks represent the selected 28 accessions.



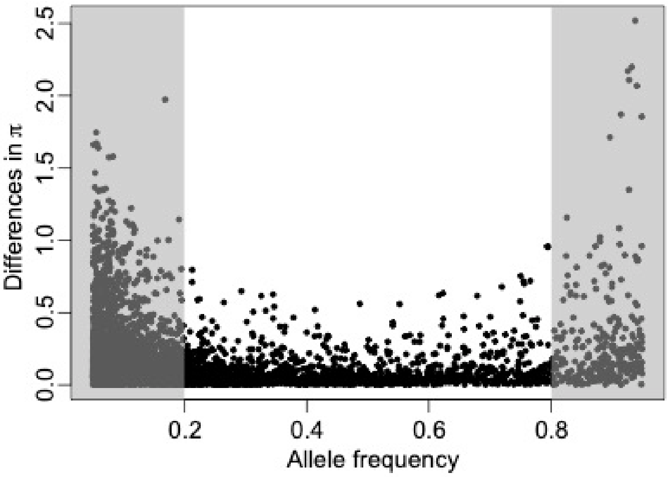
Supplemental Fig. S26.

Cross-validation error rates of ADMIXTURE analysis. The cross−validation error rates were compared at 18 models (K= 1~18). The K=13 was selected as the best model.



Supplemental Fig. S27.

Accuracy of association analysis with linear mixed model (LMM). (*A*) False positive rate of LMM at 620 accessions. (*B*) False positive rate of LMM at 28 accessions. The *X*-axis represents FDR of significance level, whereas the *Y*-axis represents false positive rate (see Methods).



Supplemental Fig. S28.

Relationship between differences in p and allele frequency of SNPs. The 4,500 SNPs were randomly selected from 1,397,934 SNPs. The differences in p were calculated from the SNPs (see Methods).