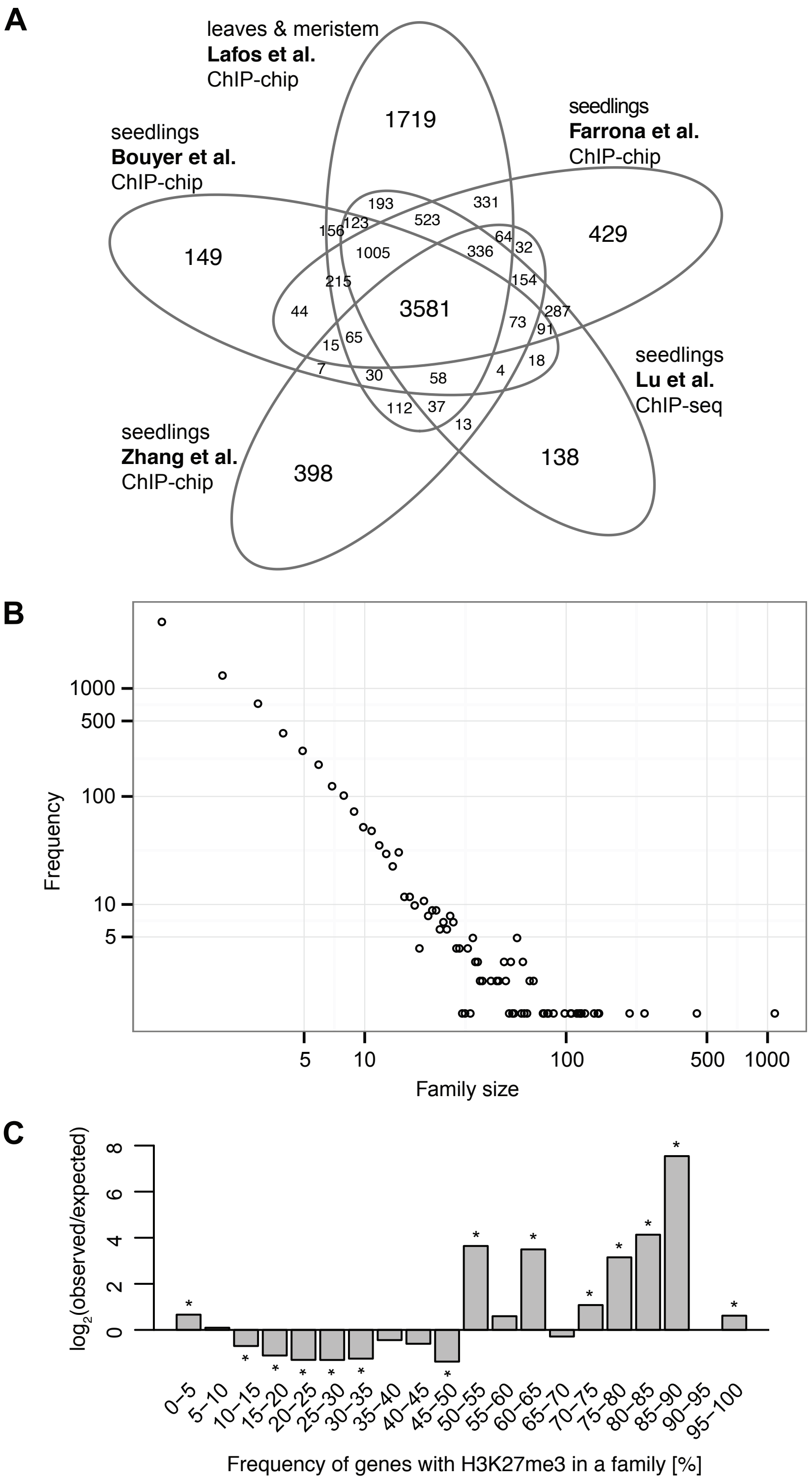
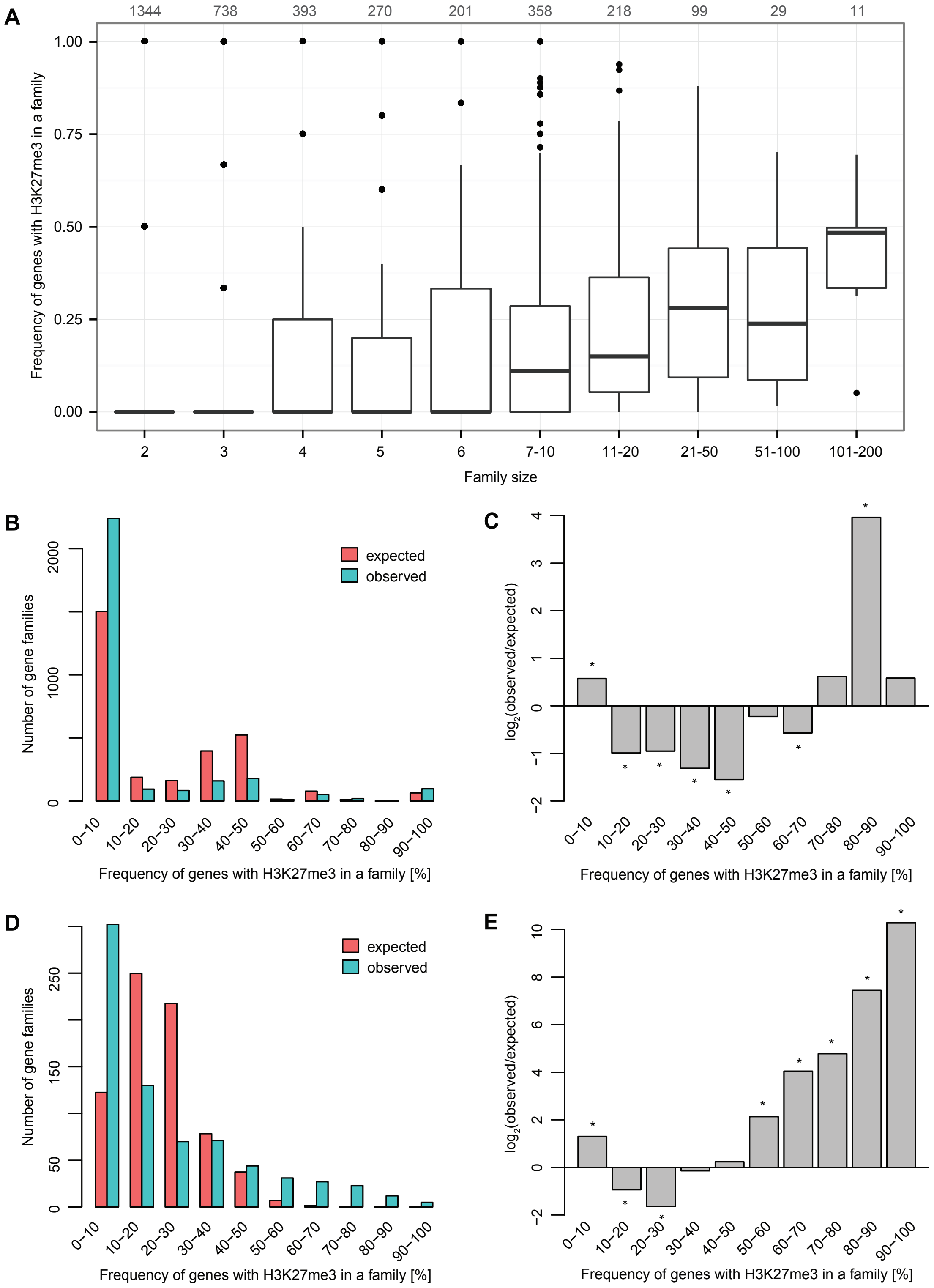
Supplementary Figures

Berke L. and Snel B.

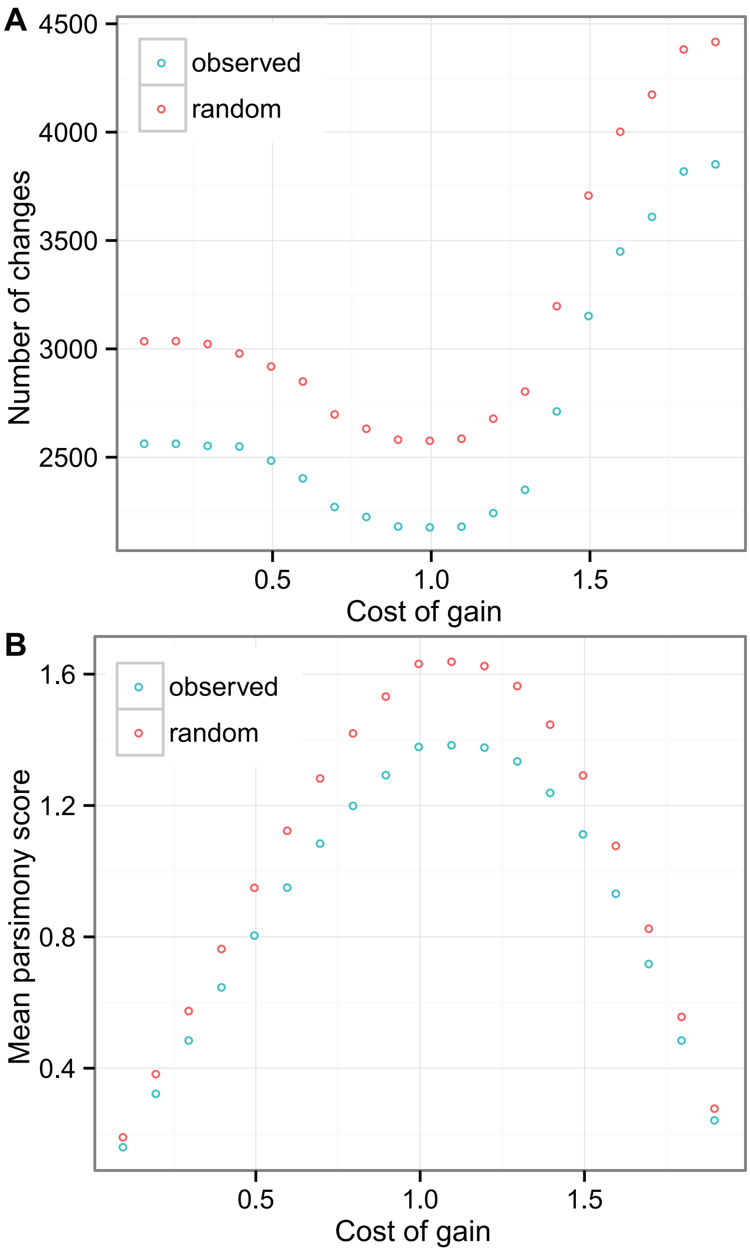
The histone modification H3K27me3 is retained after gene duplication and correlates with conserved non-coding sequences in Arabidopsis



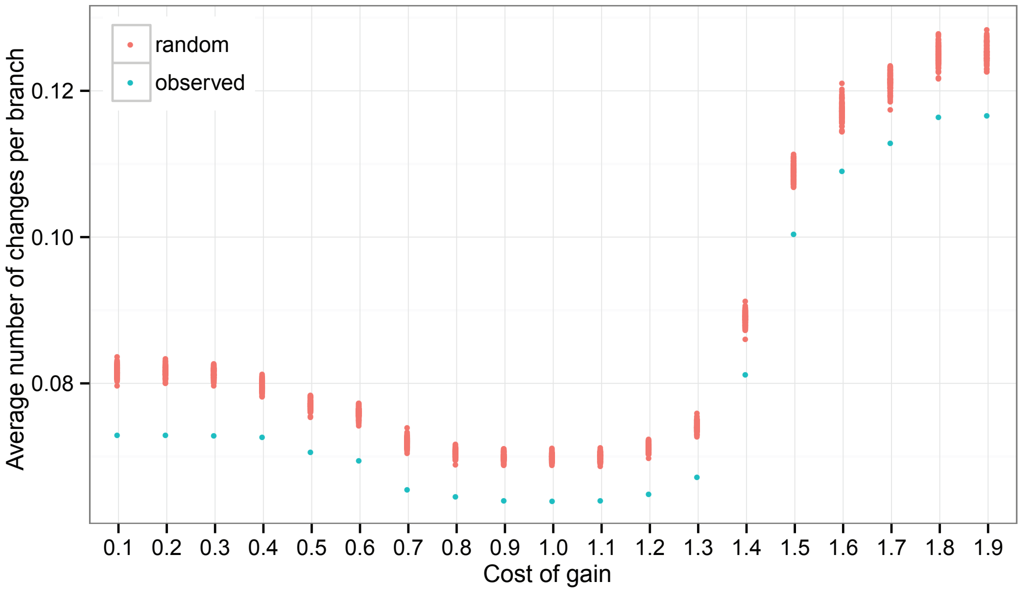
Sup. Fig. S1: **Description of datasets.** (A) H3K27me3-marked genes according to different datasets. (B) The size of constructed gene families shows exponential distribution. (C) Bar heights show an average $\log_2$ ratio of observed/expected number of gene families in each bin for 1000 permutations. Stars indicate bars where observed values were more extreme than the expected values for all 1000 permutations. The data is the same as in Fig. 1B but with smaller bin widths.



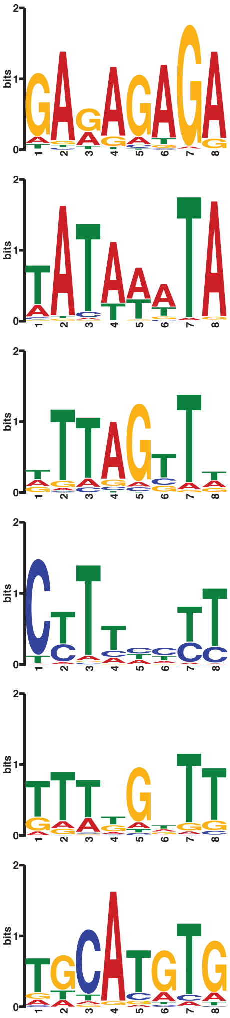
Sup. Fig. S2: **Relationship between gene family size and proportion of genes with H3K27me3.** (A) Proportion of genes with H3K27me3 in a family as a function of family size. Gray numbers on top report the number of families in each bin. Large families have higher proportion of marked genes. (B, D) Distribution of small (2-6 genes) (B) and large (>6 genes) (D) gene families according to the proportion of H3K27me3-marked genes. (C, E) Bar heights show an average $\log_2$ ratio of observed/expected number of small (C) and large (E) gene families in each bin for 1000 permutations. Stars indicate bars where observed values were more extreme than the expected values for all 1000 permutations.



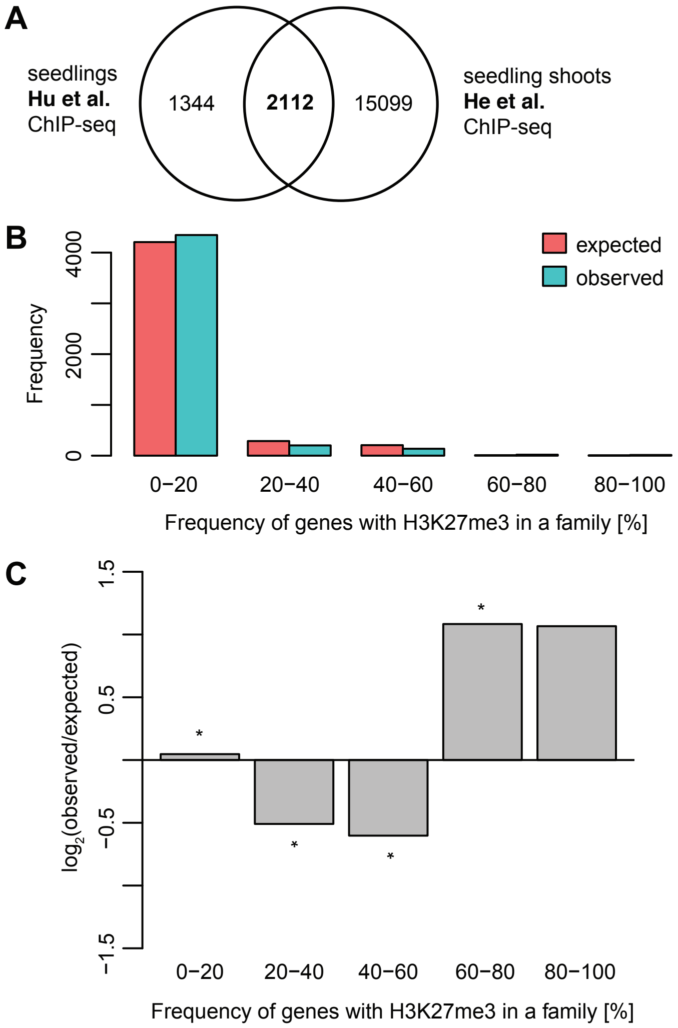
Sup. Fig. S3: **Parameter sweep for gain/loss costs.** (A) Total number of H3K27me3 changes for different gain penalties. (B) Mean parsimony score for different gain penalties. For randomized data, average value of 100 randomizations is shown.



Sup. Fig. S4: **Average rate of change of H3K27me3 per branch.** Orange data points show average number of changes per branch for each of the 100 randomizations performed.



Sup. Fig. S5: **Overrepresented motifs in CNSs associated with H3K27me3-marked genes.**



Sup. Fig. S6: **Rice gene families are enriched in either marked or non-marked genes.** (A) H3K27me3-marked genes according to different datasets. (B) Distribution of gene families according to the proportion of H3K27me3-marked genes. (C) Bar heights show an average $\log_2$ ratio of observed/expected number of gene families in each bin for 1000 permutations. Stars indicate bars where observed values were more extreme than the expected values for all 1000 permutations.