SUPPLEMENTAL METHODS

Estimation of demographic history of 620 accessions of *Arabidopsis thaliana*

Out of 1,397,934 SNPs in 620 accessions of *A*. *thaliana*, we focused on 667,277 SNPs located on the intergenic regions. To infer the demographic history in 620 *A. thaliana* accessions, we used simple five demographic models by Fastsimcoal26 program (Excoffier and Foll 2011) (Supplemental Fig. S14). Model 1 is assumed to be constant population size. Models 2 and 3 are assumed to have an event of increasing and decreasing population size, respectively. Models 4 and 5 are assumed to have an increasing event of population size after and before a decreasing event of population size, respectively. In each model, we performed Fastsimcoal26 program with the parameter '–n 100000 –d –M –L 500'. The mutation rate (μ) is defined to be 7×10^{-9} /bp/generation in *A*. *thaliana* (Ossowski et al. 2010). The likelihoods scores were compared among the five models by Akaike's information criterion score (AIC). As a result, model 5 was selected as the best fitting model

(Supplemental Table S21). This model corresponds to the demographic history reported in previous studies (Durvasula et al. 2017; The 1001 Genomes Consortium 2016). The demographic parameters estimated in model 5 are used to generate sequences under neutral model (Supplemental Table S22).

Tajima's *D* under neutral model

To generate sequences under neutral model, we performed ms program based on demographic parameters estimated by model 5 (Hudson 2002). Under ms program, sample size, replications, segment size and mutation rates are defined to be 620, 10,000, 10kb and 7×10^{-9} /bp/generations.

We converted the ms format sequences to vcf format with randomly selected SNP sites. Tajima's *D* was calculated in the converted ms sequences 10,000 replicates by VCFtools program (Danecek et al. 2011). The Tajima's *D* values in the best demographic model are expected to be neutral Tajima's *D*.

Estimation of F_{ST} and ZF_{ST}

We divided the 620 accessions into two groups, with and without mCG. The F_{ST} between the two groups was calculated in a 10-kb window by the VCFtools program (Danecek et al. 2011). To reduce the effect of population structure, we calculated *Z*-transformed F_{ST} (ZF_{ST}) from the F_{ST} values (Axelsson et al. 2013). The ZF_{ST} is estimated using the following equation: $ZF_{ST} = (F_{ST} - \mu F_{ST}) / \sigma F_{ST}$. The μF_{ST} is the mean F_{ST}. The σF_{ST} is the standard deviation. At each of selected mCGs, we calculated the ZF_{ST} in all windows. Then, we compared ZF_{ST} between windows contain selected mCG and the other windows by Wilcoxon rank sum test.

Detection of selective sweep by using SweeD

To detect selective sweep in target mCGs, we used SweeD program (Pavlidis et al. 2013). In accessions with the target mCG, we calculated composite likelihood ratio (CLR) with the parameter '–grid 2000 –maf 0.05'. We defined a significant threshold (CLR > 1.51) from top 5% of whole genome CLR. For comparison, we also calculated CLR at any genome position in accessions

lacking the focused mCG.

Supplemental References

The 1001 Genomes Consortium. 2016. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166**: 481–491.

Axelsson E, Ratnakumar A, Arendt ML, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar Å, Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* **495**: 360–364.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T, Burbano HA, Xavier Pico F, Alonso-Blanco C, et al. 2017. African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **114**: 5213–5218.

Excoffier L, Foll M. 2011. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**: 1332–1334.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in arabidopsis thaliana. *Science (80-)* **327**: 92–94.

Pavlidis P, Zivkovic D, Stamatakis A, Alachiotis N. 2013. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol Biol Evol* **30**: 2224–2234.