

Cis-meQTL mapping reveals genetic architecture of DNA methylation in a giga-genome species

Victor Chano^{1,2,*}, Konstantin V. Krutovsky^{1,2}, Kai Wang³, Matti Haapanen⁴, Fred O. Asiegbu³, Oliver Gailing^{1,2,*}

¹ Department of Forest Genetics and Forest Tree Breeding, University of Göttingen. Büsgenweg 2, 37077 Göttingen, Germany.

² Center for Integrated Breeding Research (CiBreed), University of Göttingen. Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany.

³ Department of Forest Science, University of Helsinki. FIN-00014 Helsinki, Finland.

⁴ Natural Resources Institute Finland (LUKE). Latokartanonkaari 9, Helsinki 00790, Finland.

* Corresponding authors

ORCID:

Victor Chano 0000-0002-3423-9090

Konstantin V. Krutovsky 0000-0002-8819-7084

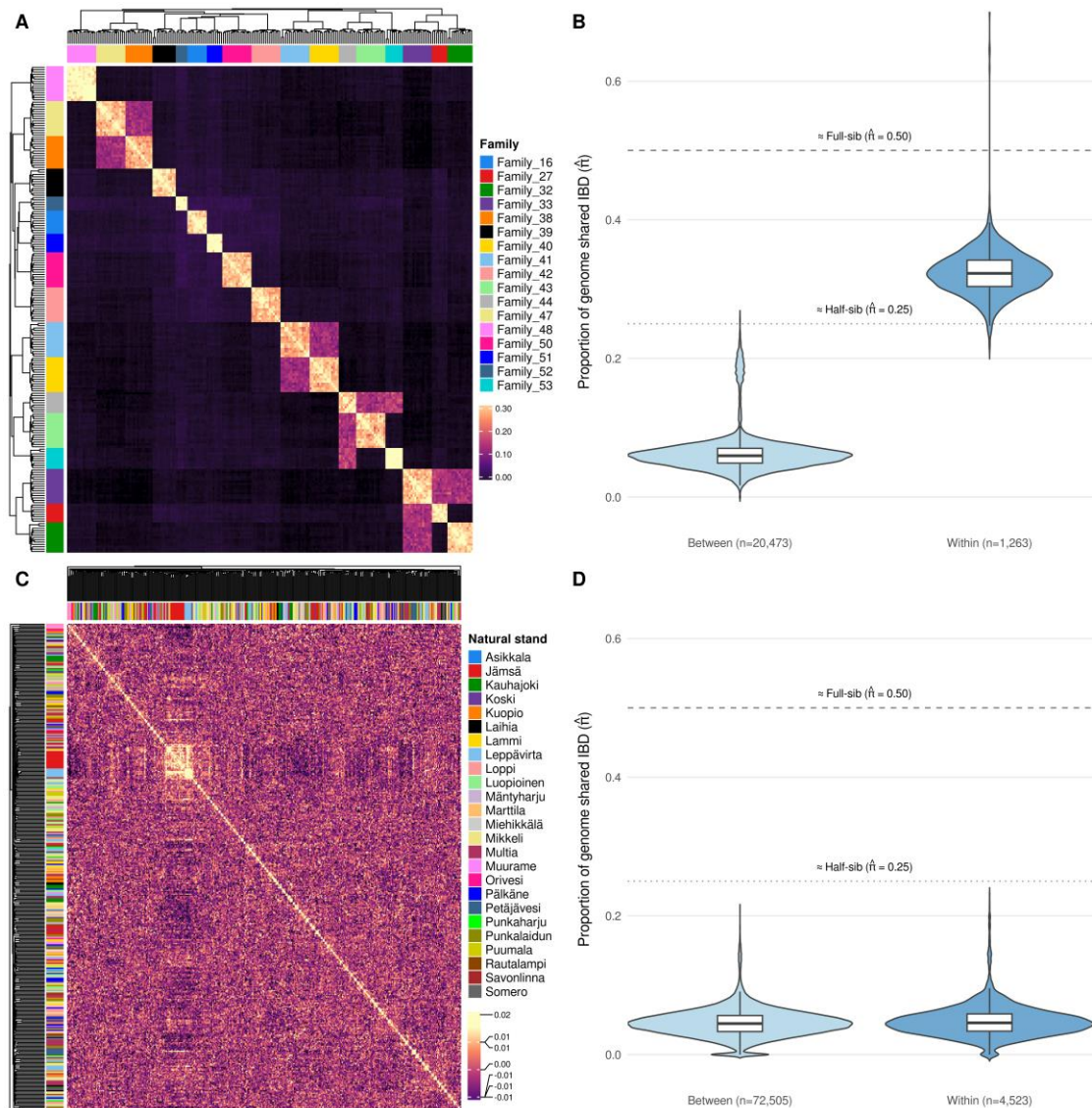
Kai Wang 0000-0002-5205-7346

Matti Haapanen 0000-0003-3294-501X

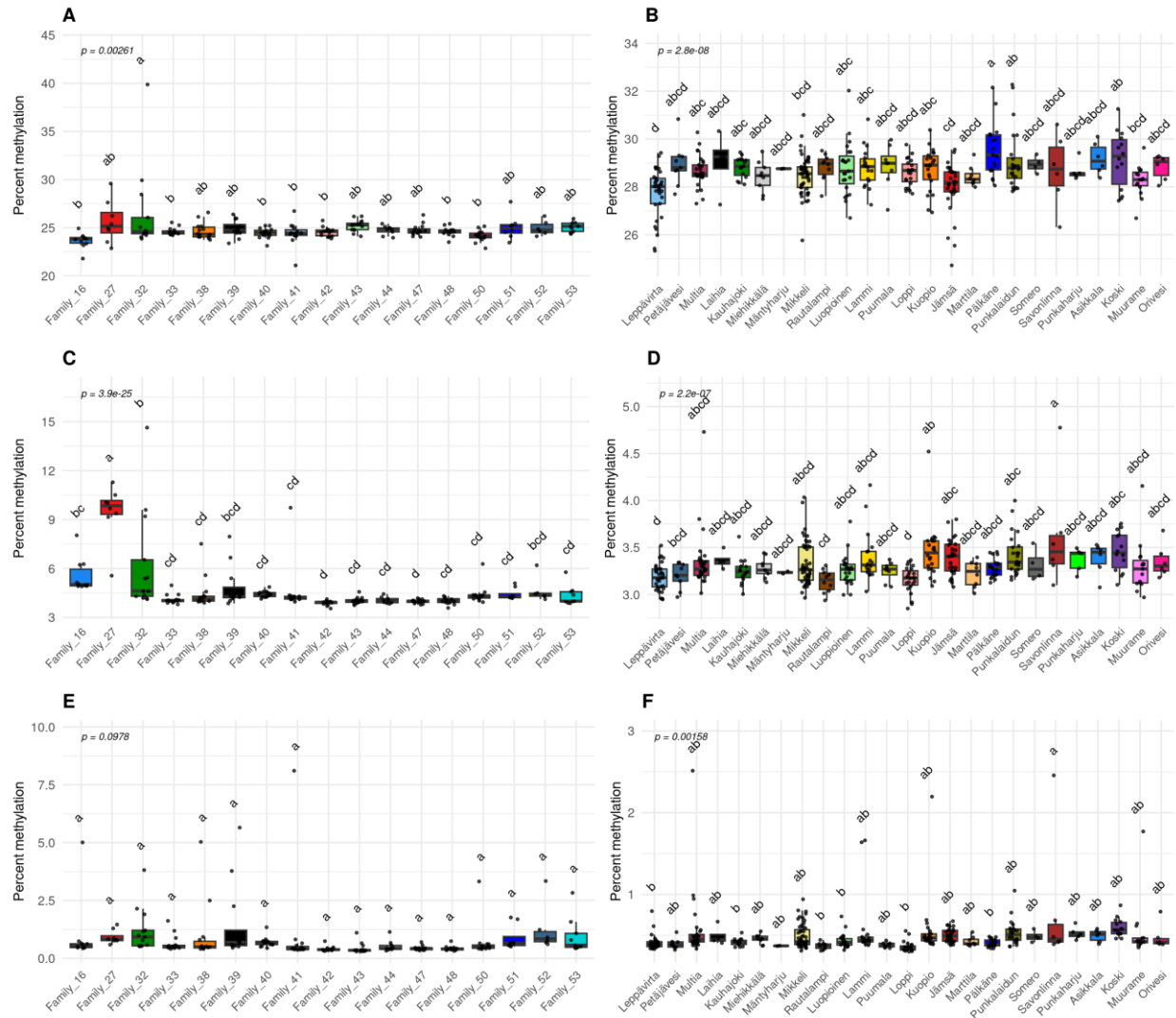
Fred O. Asiegbu 0000-0003-0223-7194

Oliver Gailing 0000-0002-4572-2408

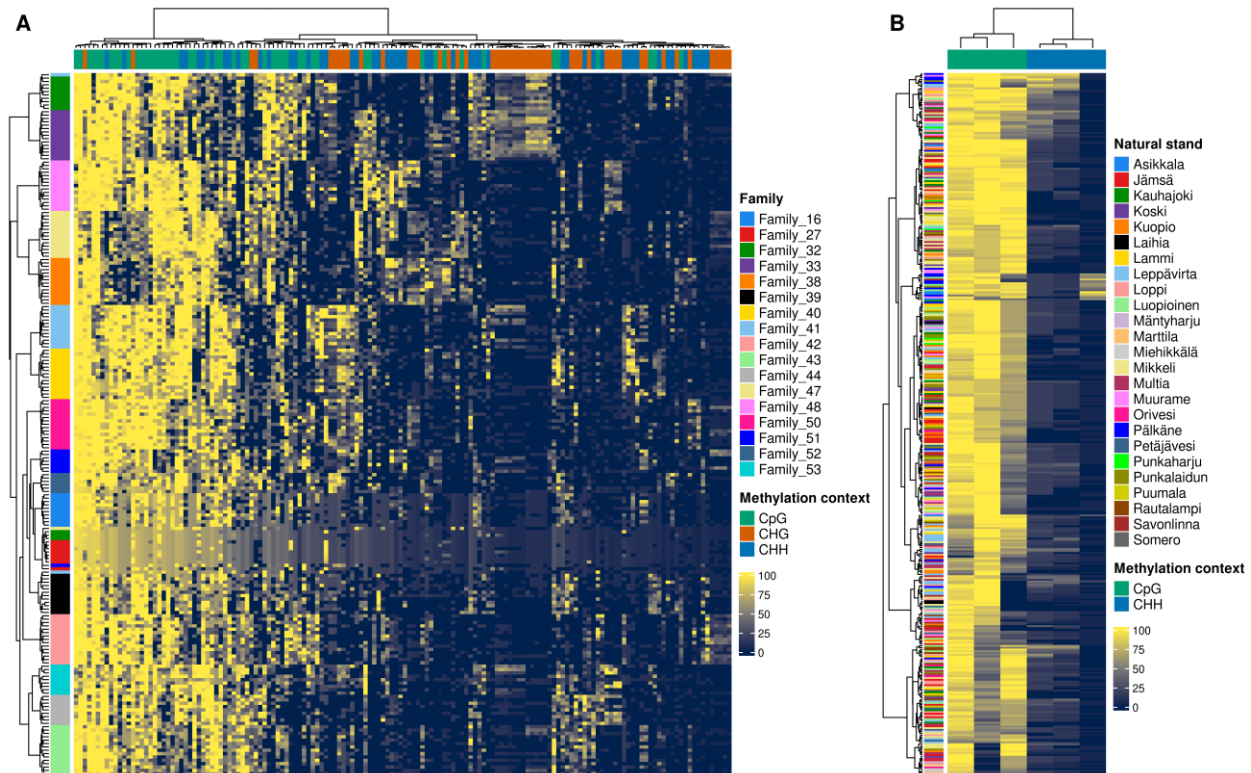
22 Extended Data



Extended Data Fig. 1 | Genomic relationship matrix (GRM) and identity-by-descent (IBD) in the breeding and natural cohorts. A–B GRM (**A**) and pairwise IBD estimates (**B**) for the breeding cohort ($n = 209$ individuals; 17 full-sib families). **C–D** GRM (**C**) and pairwise IBD estimates (**D**) for the natural cohort ($n = 393$ individuals; 25 natural stands). In **A** and **C**, rows and columns represent individual samples ordered by hierarchical clustering; colour bars indicate family (**A**) or stand (**C**) membership; cell colour encodes GRM value as indicated by the colour scale; GRM was estimated from 450,947 SNPs (**A**) and 462,617 SNPs (**C**). In **A**, relatedness within family blocks spans approximately 0.10 to over 0.30, and cross-family relatedness is apparent between Families 38 and 47, between Families 40 and 41, and between Family 33 and Families 32 and 27. In **C**, off-diagonal values are uniformly low on a compressed scale (-0.01 to 0.02). In **B** and **D**, violin plots show pairwise $\hat{\pi}$ estimates for between-family ($n = 20,473$) and within-family ($n = 1,263$) pairs (**B**) and between-stand ($n = 72,505$) and within-stand ($n = 4,523$) pairs (**D**); dashed and dotted reference lines indicate the expected $\hat{\pi}$ for full-sib ($\hat{\pi} = 0.50$) and half-sib ($\hat{\pi} = 0.25$) relationships, respectively. SNP, single-nucleotide polymorphism; $\hat{\pi}$, proportion of the genome shared IBD.

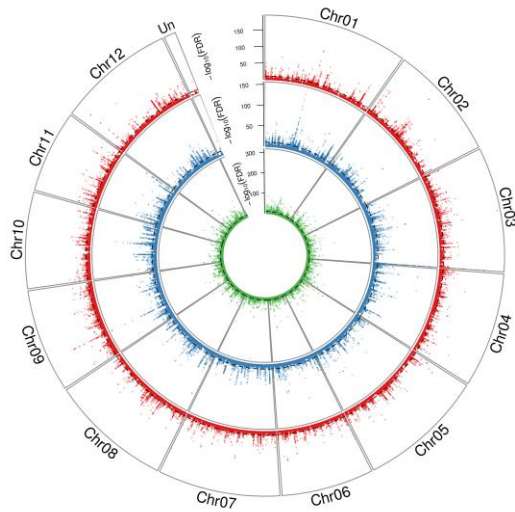


Extended Data Fig. 2 | Global methylation levels per family and natural stand across three methylation contexts. **A–B** Boxplots of per-sample mean CpG methylation for the breeding (**A**; $n = 209$ individuals; 17 full-sib families) and natural (**B**; $n = 393$ individuals; 25 natural stands) cohorts. **C–D** Boxplots of per-sample mean CHG methylation for the breeding (**C**) and natural (**D**) cohorts. **E–F** Boxplots of per-sample mean CHH methylation for the breeding (**E**) and natural (**F**) cohorts. In all panels, boxes show the median and interquartile range; individual points represent per-sample values. One-way ANOVA global p -values are displayed in each panel. Lowercase letters above each box depict significance of Tukey's HSD post-hoc pairwise comparisons (adjusted $p < 0.05$); boxes sharing a letter are not significantly different. CpG, CHG, and CHH, cytosine methylation contexts (H = A, T or C).

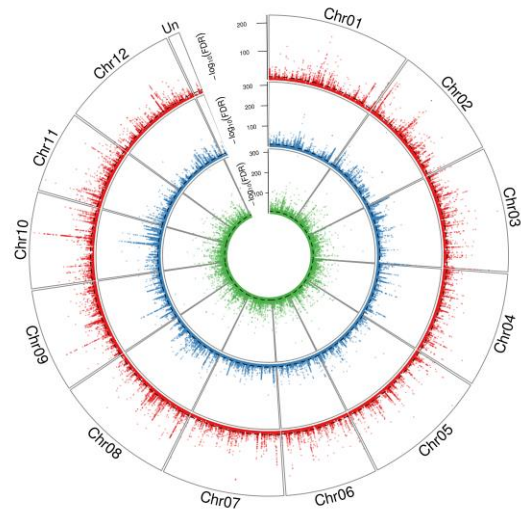


Extended Data Fig. 3 | Hierarchically clustered heatmaps of significantly variable methylated positions (SVMPs). **A** Heatmap for the breeding cohort ($n = 209$ individuals; 17 full-sib families) based on the top-150 SVMPs (50 highest-ranking loci per methylation context by adjusted p -value). **B** Heatmap for the natural cohort ($n = 393$ individuals; 25 natural stands) based on all 6 significant SVMPs (adjusted p -value < 0.05 ; 3 CpG and 3 CHH loci). In both panels, rows represent individual samples and columns represent methylation loci; rows and columns are independently ordered by hierarchical clustering (Euclidean distance, Ward linkage). Colour bars on the left indicate family (**A**) or natural stand (**B**); colour bars on top indicate methylation context. Cell colour encodes methylation percentage as indicated by the colour scale. Missing values were imputed by locus mean exclusively for hierarchical clustering; original unimputed values were used for all statistical tests. SVMPs were identified by the Kruskal-Wallis H-test applied independently per locus, with Benjamini–Hochberg (BH) correction within each methylation context (adjusted p -value < 0.05). CpG, CHG, and CHH, cytosine methylation contexts (H = A, T or C).

A



B



● CpG ● CHG ● CHH

Extended Data Fig. 4 | Genome-wide circular Manhattan plots of cis-meQTL associations (MatrixEQTL). **A** Breeding cohort (n = 209 individuals; 17 full-sib families). **B** Natural cohort (n = 393 individuals; 25 natural stands). Each point represents one tested SNP–methylation site pair within a 100 kb cis window; the dashed line marks the significance threshold ($FDR < 1 \times 10^{-10}$), and only points beyond it are significant; points are coloured in red for CpG, blue for CHG, and green for CHH methylation contexts. Chromosomes Chr01–Chr12 are arranged clockwise from the top, followed by the sector labelled "Un", which holds unplaced contigs; radial distance is proportional to $-\log_{10}(FDR)$; radial scales differ between rings and between panels. Genomic inflation factors (λ) per context are presented in Table 1. FDR, false discovery rate; CpG, CHG, and CHH, cytosine methylation contexts (H = A, T or C).