Natural *CMT2* variation is associated with genome-wide methylation changes and temperature adaptation

Xia Shen[1], Jennifer De Jonge[2], Simon Forsberg[1], Mats Pettersson[1], Zheya Sheng[1], Lars Hennig[2] and Örjan Carlborg[1]*

Affiliations: [1]Swedish University of Agricultural Sciences, Department of Clinical Sciences, Division of Computational Genetics, Box 7078, SE-750 07 Uppsala, Sweden. [2]Swedish University of Agricultural Sciences, Department of Plant Biology, Box 7080, SE-750 07 Uppsala, Sweden.

*To whom correspondence should be addressed. Email: orjan.carlborg@slu.se

ABSTRACT

As *Arabidopsis thaliana* has colonized a wide range of habitats across the world it is an attractive model for studying the genetic mechanisms underlying environmental adaptation. Here, we used public data from two collections of *A. thaliana* accessions to associate genetic variability at individual loci with differences in climates at the sampling sites. We use a novel method to screen the genome for plastic alleles that tolerate a broader climate range than the major allele. This approach reduces confounding with population structure and increases power compared to standard genome-wide association methods. Sixteen novel loci were found, including an association between Chromomethylase 2 (*CMT2*) and variability in seasonal temperatures where the plastic allele had reduced genome-wide CHH methylation. *Cmt2* mutants were more tolerant to heat-stress, suggesting genetic regulation of epigenetic modifications as a likely mechanism underlying natural adaptation to variable temperatures, potentially through differential allelic plasticity to temperature-stress.

AUTHOR SUMMARY

A central problem when studying adaptation to a new environment is the interplay between genetic variation and phenotypic plasticity. *Arabidopsis thaliana* has colonized a wide range of habitats across the world and it is therefore an attractive model for studying the genetic mechanisms underlying environmental adaptation. Here, we used publicly available data from two collections of *A. thaliana* accessions, covering the native range of the species, to identify loci associated with differences in climates at the sampling sites. To address the confounding between geographic location, climate and population structure, a new genome-wide association analysis method was developed that facilitates detection of potentially adaptive loci where the alternative alleles display

different tolerable climate ranges. Sixteen novel such loci, many of which contained candidate genes with amino acid changes, were found including a strong association between Chromomethylase 2 (*CMT2*) and variability in seasonal temperatures. The reference allele dominated in areas with less seasonal variability in temperature, and the alternative allele, which disrupts genome-wide CHH-methylation, existed in both stable and variable regions. We have experimentally shown that plants with a defective *CMT2* gene tolerate heat-stress better than plants with a functional gene. Our results thus link natural variation in *CMT2*, and differential genome-wide CHH methylation, to the distribution of *A. thaliana* accessions across habitats with different seasonal temperature variability and show that disruption of *CMT2* function improves heat-stress tolerance. The results therefore also suggest a role for genetic regulation of epigenetic modifications in natural adaptation to temperature, potentially through differential allelic plasticity, and illustrate the importance of re-analyses of existing data using new analytical methods to obtain a more complete understanding of the mechanisms contributing to adaptation.

BLURB

Disrupted Chromomethylase 2 (*CMT2)* is associated with plasticity to seasonal temperature variability, decreased CHH-methylation and improved heat-stress tolerance, suggesting a role for genetic regulation of epigenetic modifications in natural adaptation.

INTRODUCTION

Arabidopsis thaliana has colonized a wide range of habitats across the world and it is therefore an attractive model for studying the genetic mechanisms underlying environmental adaptation [1]. Several large collections of *A. thaliana* accessions have either been whole-genome re-sequenced or high-density SNP genotyped [1-7]. The included accessions have adapted to a wide range of different climatic conditions and therefore loci involved in climate adaptation will display genotype by climate-at-sampling-site correlations in these populations. Genome-wide association or selective-sweep analyses can therefore potentially identify signals of natural selection involved in environmental adaptation, if those can be disentangled from the effects of other population genetic forces acting to change the allele frequencies. Selective-sweep studies are inherently sensitive to population-structure and, if present, the false-positive rates will be high as the available statistical methods are unable to handle this situation properly. Further experimental validation of inferred sweeps (e.g. [1,8]) is hence necessary to suggest them as adaptive. In GWAS, kinship correction is

now a standard approach to account for population structure that properly controls the false discovery rate. Unfortunately, correcting for genomic kinship often decreases the power to detect individual adaptive loci, which is likely the reason that no genome-wide significant associations to climate conditions were found in earlier GWAS analyses [1,8]. Nevertheless, a number of candidate adaptive loci could despite this be identified using extensive experimental validation [1,2,8], showing how valuable these populations are as a resource for finding the genomic footprint of climate adaptation.

RESULTS

Genome-wide association analysis to detect loci with plastic response to climate

Here, we re-analyze the data from the RegMap collection to find loci contributing to climate adaptation through an alternative mechanism: genetic control of plasticity. Such loci are unlikely to be detected with standard GWAS or selective-sweep analyses as they have a different genomic signature of selection and distribution across climate envelopes [9]. We extend and utilize an approach [10,11] that instead of mapping loci by differences in allele-frequencies between local environments, which is highly confounded by population structure, infer adaptive loci using a heterogeneity-of-variance test. This identifies loci where the minor allele tolerate a broader range of climate conditions than the major allele [10]. As such plastic alleles will be present across the entire population, they are less confounded with population structure and detectable in our GWAS analysis that utilizes kinship correction to account for population stratification. A genome-wide association analysis was performed for thirteen climate variables across ~215,000 SNPs in 948 *A. thaliana* accessions from the RegMap collection, representing the native range of the species [1,12]. In total, sixteen genome-wide significant loci were associated with eight climate variables (Table 1), none of which could be found using standard methods for GWAS analyses [1,8,13-15]. The effects were in general quite large, from 0.3 to 0.5 residual standard deviations (Table 1), corresponding to increases in the climate plasticity of 21-35% from the alternate allele. The detailed results for each trait are reported in Text S1 (Supplementary Figure 1-13). The distributions of the significant plastic alleles across the population strata in relation to their geographic origin and the climate envelopes are provided in Text S1 (Supplementary Figure 14-35).

Fine-mapping and identification of candidate mutations in the 1001-genomes data

Utilizing data from the 1001-genomes project [2-7] (http://1001genomes.org), we fine-mapped the significant loci and identified five functional candidate genes (Table 1) and 11 less well

characterized genes (Table S1) with either missense, nonsense or frameshift mutations in high linkage disequilibrium (LD; $r^2 > 0.8$) with the leading SNP of each locus. 76 additional linked loci or genes without candidate mutations in the coding regions are reported in Table S2.

Chromomethylase 2 (*CMT2*) is associated with seasonal temperature variability in the RegMap collection

A strong association to temperature seasonality, i.e. the ratio between the standard deviation and the mean of temperature records over a year, was identified near Chromomethylase 2 (*CMT2*; Table 1; Figure 1). Stable areas are generally found near large bodies of water (e.g. London near the Atlantic 11 ± 5◦C; mean ± SD) and variable areas inland (e.g. Novosibirsk in Siberia 1 ± 14◦C). A premature *CMT2* stop codon located on chromosome 4 at 10 414 556 bp segregated in the RegMap collection This $CMT2_{STOP}$ allele had a genome-wide significant association to temperature seasonality ($P = 1.1\times10^{-7}$) and was in strong LD ($r^2 = 0.82$) with the leading SNP (Figure 1B). The geographic distributions of the wild-type ($CMT2_{WT}$) and the alternative ($CMT2_{STOP}$) alleles in the RegMap collection shows that the $CMT2_{WT}$ allele dominates in all major sub-populations sampled from areas with low or intermediate temperature seasonality. The plastic $CMT2_{STOP}$ allele is present, albeit at lower frequency, across all sub-populations in low- and intermediate temperature seasonality areas, and is more common in areas with high temperature seasonality (Figure 2A; Figure 3; Text S1, Supplementary Figure 36). Such global distribution across the major population strata indicates that the allele has been around in the Euroasian population sufficiently long to spread across most of the native range and that lack of functional *CMT2* is not deleterious but rather maintained through balancing selection [9], perhaps through an improved tolerance to variable temperatures.

Broader geographic distribution of the $CMT2_{STOP}$ allele in the 1001-genomes collection

To confirm that the $CMT2_{STOP}$ association was not due to sampling bias in the RegMap, we also scored the *CMT2* genotype and collected the geographical origins from 665 accessions in the 1001-genomes project (http://1001genomes.org) [2,3,5-7]. In this more geographically diverse set (Figure 2A), $CMT2_{STOP}$ was more common (MAF = 0.10) and had a similar allele distribution across Euroasia as in RegMap (Text S1, Supplementary Figure 36-37). Two additional mutations were identified on unique haplotypes ($r2 = 0.00$) - one nonsense $CMT2_{STOP2}$ at 10 416 213 bp (MAF = 0.02) and a frameshift mutation at 10 414 640 bp (two accessions). Both $CMT2_{STOP}$ and $CMT2_{STOP2}$ had genotype-phenotype maps implying a plastic response to variable temperature (Figure 2B) and

the existence of multiple mutations disrupting *CMT2* further suggest lack of *CMT2* function as a potentially evolutionary beneficial event.

Accessions with the $CMT2_{STOP}$ allele has an altered genome-wide CHH-methylation pattern

CMT2 is a plant DNA methyltransferase that methylates mainly cytosines in CHH (H = any base but G) contexts, predominantly at transposable elements (TEs) [16,17]. We tested the effect of $CMT2_{STOP}$ on genome-wide DNA methylation using 131 $CMT2_{WT}$ and 17 $CMT2_{STOP}$ accessions, for which MethylC-sequencing data was publicly available [7]. A methylome-wide association (MWA) analysis between $CMT2_{STOP}$ and the methylation-state at ~6 million single methylation polymorphisms (SMPs) identified ~3,000 methylome-wide significant CHH associations (Text S1, Supplementary Figure 38). Methylation was homogenous across the clear majority of the CHH-sites for the $CMT2_{WT}$ accessions, but not for CG or CHG-sites (Text S1, Supplementary Figure 39). Interestingly, the methylation-pattern is more heterogenous among the $CMT2_{STOP}$ accessions, with a small overlap both among $CMT2_{STOP}$ accessions and between $CMT2_{STOP}$ and $CMT2_{WT}$ accessions, indicating a shared residual, non-*CMT2* mediated CHH methylation. We confirmed that the differential methylation detected in the $CMT2_{STOP}$ and $CMT2_{WT}$ accessions is consistent with the effects of disrupting *CMT2*, by showing that the level of CHH-methylation across the MWA detected sites was significantly lower ($P = 2.2 \times 10^{-4}$) in four T-DNA insertion *cmt2* knock-out alleles [16] than in $CMT2_{WT}$ plants (Figure 4A; Text S1, Supplementary Figure 39B).

***Cmt2* mutant plants have an improved heat-stress tolerance**

To functionally explore whether lack of *CMT2* alters the response to temperature-stress, we subjected Col-0 and two independent *cmt2* mutants to heat-stress. The *cmt2* mutants had significantly higher survival-rate (2.1-fold; $P = 2.7 \times 10^{-3}$; Figure 4B) than Col-0. This striking improvement in tolerance to heat-stress of *cmt2* plants suggests *CMT2*-dependent CHH methylation as an important alleviator of stress responses in A. thaliana and a candidate mechanism for temperature adaptation.

DISCUSSION

A major challenge in attempts to identify individual loci involved in climate adaptation is the strong confounding between geographic location, climate and population structure in the natural A. thaliana population. Earlier genome-wide association analyses in large collections of natural accessions experienced a lack of statistical power when correcting for population-structure [1,8].

We used an alternative GWAS approach [10] to test for a variance-heterogeneity, instead of a mean difference, between genotypes. This analysis identifies loci where the minor allele is more plastic (i.e. exist across a broader climatic range) than the major allele. In contrast, a standard GWAS map loci where the allele-frequencies follow the climatic cline. Although plastic alleles might be less frequent in the genome, they are easier to detect in this data due to their lower confounding with population-structure. This overall increase in power is also apparent when comparing the signals that reach a lower, sub-GWAS significance level (Text S1, Supplementary Figure 40-44). In the *CMT2*-locus, that was strongly associated with variable seasonal temperatures, the plastic allele had a disrupted genome-wide CHH methylation pattern similar to that of *cmt2* mutant plants. Interestingly, we could also show that *cmt2* mutants had an increased tolerance to heat-stress, a finding that both strongly implicates *CMT2* as an adaptive locus and clearly illustrates the potential of our method as a useful approach to identify novel associations of functional importance.

It is not clear via which mechanism *CMT2*-dependent CHH methylation affects plant heat tolerance. We consider it most likely that the effect will be mediated by TEs in the immediate neighborhood of protein-coding genes. Heterochromatic states at TEs can affect activity of nearby genes and thus potentially plant fitness [18]. Consistent with a repressive role of *CMT2* on heat stress responses, *CMT2* expression is reduced by several abiotic stresses including heat [19]. Because global depletion of methylation has been shown to enhance resistance to biotic stress [20], it is possible that DNA-methylation has a broader function in shaping stress responses than currently thought.

We identified several alleles associated with a broader range of climates across the native range of A. thaliana, suggesting that a genetically mediated plastic response might of important for climate adaptation. A defective epigenetic mechanism involving *CMT2* mediated CHH-methylation was strongly associated with adaptation to variability in seasonal temperatures and *cmt2* plants tolerated heat-stress better than wild-type plants. Together, these findings suggest genetically determined epigenetic variability as a likely mechanism contributing to phenotypic plasticity of adaptive advantage in natural environments.

MATERIALS AND METHODS

Climate data and genotyped *Arabidopsis thaliana* accessions

The climate phenotypes and *A. thaliana* genotype data that we re-analyzed were obtained from [1]. 13 climate variables and genotypes of 214,553 single nucleotide polymorphisms (SNPs) for 948

accessions were available at: http://bergelson.uchicago.edu/regmap-data/climate-genome-scan. The climate variables used in the analyses were: aridity, number of consecutive cold days (below 4 degrees Celsius), number of consecutive frost-free days, day-length in the spring, growing-season length, maximum temperature in the warmest month, minimum temperature in the coldest month, temperature-seasonality, photosynthetically active radiation, precipitation in the wettest month, precipitation in the driest month, precipitation-seasonality, and relative humidity in the spring. No squared pairwise Pearson's correlation coefficients between the phenotypes were greater than 0.8 (Figure S7 of [1]).

Phenotyping of temperature seasonality in Euro-Asia for the 1001-genomes (http://1001genomes.org) accessions was done by downloading the raw data from http://www.worldclim.org/. The data were re-formatted and thereafter processed by the raster package in R. The R code is provided in the Text S1.

Genome-wide association analysis to identify adaptability loci in the RegMap collection

Genome-wide association (GWA) datasets based on natural collections of *A. thaliana* accessions, such as the RegMap collection, are often genetically stratified. This is primarily due to the close relationships between accessions sampled at nearby locations. Furthermore, as the climate measurements used as phenotypes for the accessions are values representative for the sampling locations of the individual accessions, these measurements will be confounded with the general genetic relationship [12]. Unless properly controlled for, this confounding might lead to excessive false-positive signals in the association analysis; this as the differences in allele-frequencies between loci in locations that differ in climate, and at the same time are geographically distant, will create an association between the genotype and the trait. However, this association could also be due to other forces than selection. In traditional GWA analyses, mixed-model based approaches are commonly used to control for population-stratification. The downside of this approach is that it, in practice, will remove many true genetic signals coming from local adaptation due to the inherent confounding between local genotype and adaptive phenotype. Instead, the primary signals from such analyses will be due to effects of alleles that exist in, and have similar effects across, the entire studied population. In general, studies into the contributions of genetic variance-heterogeneity to the phenotypic variability in complex traits is a novel and useful approach with great potential [21]. Here, we have developed and used a new approach that combines a linear mixed model and a variance-heterogeneity test, which addresses these initial concerns and shown that it is possible to

infer statistically robust results of genetically regulated phenotypic variability in GWA data from natural populations.

Statistical modeling in genome-wide scans for adaptability

The climate data at the geographical origins of the *A. thaliana* accessions were treated as phenotypic responses. Each climate phenotype vector **y** for all the accessions were normalized via an inverse-Gaussian transformation. The squared normalized measurement $z_i = y_i^2$ of accession i is modeled by the following linear mixed model to test for an association with climate adaptability (i.e. a greater plasticity to the range of the environmental condition):

$$\mathrm{z}_i = \mu + \beta x_i + g_i + e_i$$

where μ is an intercept, x_i the SNP genotype for accession i, β the genetic SNP effect, $\mathbf{g} \sim MVN(\mathbf{0}, \mathbf{G}^* \sigma_g^2)$ the polygenic effects and $\mathbf{e} \sim MVN(\mathbf{0}, \mathbf{I}\sigma_e^2)$ the residuals. x_i is coded 0 and 2 for the two homozygotes (inbred lines). The genomic kinship matrix $\mathbf{G}^*$ is constructed via the whole-genome generalized ridge regression method HEM (heteroscedastic effects model) [11] as $\mathbf{G}^* = \mathbf{ZWZ}'$,where $\mathbf{Z}$ is a number of individuals by number of SNPs matrix of genotypes standardized by the allele frequencies. $\mathbf{W}$ is a diagonal matrix with element $w_{jj} = \hat{b}_j/(1 - h_{jj})$ for the j-th SNP, where $\hat{b}_j$ is the SNP-BLUP (SNP Best Linear Unbiased Prediction) effect estimate for the j-th SNP from a whole-genome ridge regression, and h_{jj} is the hat-value for the j-th SNP. Quantities in $\mathbf{W}$ can be directly calculated using the $\mathbf{bigRR}$ package [11] in R. An example R source code for performing the analysis is provided in the Text S1.

The advantage of using the HEM genomic kinship matrix $\mathbf{G}^*$, rather than an ordinary genomic kinship matrix $\mathbf{G} = \mathbf{ZZ}'$, is that HEM is a significant improvement of the ridge regression (SNP-BLUP) in terms of the estimation of genetic effects [11,22]. Due to this, the updated genomic kinship matrix $\mathbf{G}^*$ better represents the relatedness between accessions and also accounts for the genetic effects of the SNPs on the phenotype.

Testing and quality control for association with climate adaptability

The test statistic for the SNP effect β is constructed as the score statistic [23]:

$$T^2 = \frac{(\tilde{\mathbf{x}}'\mathbf{G}^{*-1}\tilde{\mathbf{z}})^2}{\tilde{\mathbf{x}}'\mathbf{G}^{*-1}\tilde{\mathbf{x}}}$$

implemented in the GenABEL package [24], where $\tilde{\mathbf{x}} = \mathbf{x} - E[\mathbf{x}]$ are the centered genotypic values and $\tilde{\mathbf{z}} = \mathbf{z} - E[\mathbf{z}]$ the centered phenotypic measurements. The T^2 statistic has an asymptotic χ^2 distribution with 1 degree of freedom. Subsequent genomic control (GC) [25] of the genome-wide association results was performed under the null hypothesis that no SNP has an effect on the climate phenotype. SNPs with minor allele frequency (MAF) less than 0.05 were excluded from the analysis. A 5% Bonferroni-corrected significance threshold was applied. As suggested by [26], the significant SNPs were also analyzed using a Gamma generalized linear model to exclude positive findings that might be due to low allele frequencies of the high-variance SNP.

Functional analysis of polymorphisms in loci with significant genome-wide associations to climate

All the loci that showed genome-wide significance in the association study was further characterized using the genome sequences of 728 accessions sequenced as part of the 1001-genomes project (http://1001genomes.org). Mutations within a ±100Kb interval of each leading SNP, and that are in high linkage disequilibrium (LD) with the leading SNP ($r^2 > 0.8$), were reported (Table S1). The consequences of the identified polymorphisms were predicted using the Ensembl variant effect predictor [27] and their putative effects on the resulting protein estimated using the PASE (Prediction of Amino acid Substitution Effects) tool [28].

Methylome-wide association analysis and validation of *CMT2$_{STOP}$* genotypes

A methylome-wide association (MWA) analysis was conducted to the *CMT2$_{STOP}$* genotypes at 43,182,344 scored single methylation polymorphisms (SMPs) across the genome [7]. 131 *CMT2$_{WT}$* and 17 *CMT2$_{STOP}$* accessions, for which MethylC-sequencing data was publicly available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43857, were included in the analysis. Sites that were methylated at a frequency less than 0.05 among the accessions were removed from the analysis, resulting in 6,120,869 methylation sites to be tested across the genome. In this set, we tested for an association between the *CMT2$_{STOP}$* genotype and the methylation state at each of the 6,120,869 SMPs using the `qtscore()` function in the GenABEL package.

In total, 3,096 methylome-wide SMPs were significant at a Bonferroni-corrected significance threshold for 6,120,869 tests. Divided according to the type of methylated sites, they corresponded to 879 CHH, 1162 CG and 731 CHG SMPs. To visualize the pairwise similarity between accessions at these sites, we computed an identity-by-methylation-state (IBMS) matrix using all the significant CHH sites (Text S1, Supplementary Figure 39A). To validate that the high degree of shared methylated sites was a useful predictor for the $CMT2_{STOP}$ genotype, we downloaded data from an independent experiment [16] that contained methylome data on four *CMT2* knockouts and four WT samples (GSM1083504, GSM1083505, GSM1083506, GSM1014134, GSM1014135, GSM1093622 and GSM1093629 from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41302). In total, the data from [16] contained scores for 718/213/262 of the differentially methylated CHH/CHG/CG sites. The level of CHH/CHG/CG methylation was scored in each of the eight samples as the sum of the methylation levels across all these CHH/CHG/CG sites. The respective methylation-levels for all samples are provided in Text S1, Supplementary Figure 39B/C/D and a t-test shows that the methylation-level was significantly different between lines with dysfunctional and WT *CMT2* for CHH sites, but not for CHG or CG sites. Together these results clearly shows that the $CMT2_{STOP}$ accessions carry a mutated *CMT2* allele.

Our IBMS results (Text S1, Supplementary Figure 39A) indicated that four of the 17 $CMT2_{STOP}$ accessions (En-D, Fi-0, Stw-0 & Vind-1) displayed a CHH-methylation pattern across the differentially methylated sites that was closer to the $CMT2_{WT}$ phenotype. Interestingly, an evaluation of the *CMT2* mRNA abundance in these accessions using data from [7] showed that these lines also had higher transcript levels than other $CMT2_{STOP}$ accessions. Although $CMT2_{STOP}$ is a strong candidate mutation to cause the mutant *CMT2* methylation phenotype, it is not possible to rule out that it is only in strong LD with an alternative causative variant or that other mechanisms are involved that causes the obligatory epialleles across the genome to be reverted by other compensatory mechanisms. Regardless, the results clearly show that the differential CHH-methylation phenotype is caused by a loss-of-function *CMT2* allele.

Heat-stress treatments on Col-0 and *cmt2* knockouts

Seeds of Col-0, *cmt2-5* (SAIL_906_G03) and *cmt2-7* (WiscDSLox471G8) were plated on 1/2 MS medium (0.8% agar, 1% sucrose), stratified for two days at 4°C in the dark and transferred to a growth chamber with 16h light (110 µmol m^{-2} s^{-1}, 22°C) and 8h dark (20°C) periods. Ten-day old

seedlings were transferred to 4°C for one hour and subsequently placed for 24h at 37.5°C in the dark. Plant survival was scored two days after heat stress.

No difference in survival rate was found between the two knockouts *cmt2-5* and *cmt2-7*. A log-linear regression was conducted to test for the difference in survival rate between Col-0 and *cmt2* knockouts, i.e.

$$\log\left(E\left[\frac{s_i}{t_i}\right]\right) = \beta_0 + E_i + a_i$$

where s_i is the number of survived plants of accession i, t_i the corresponding total number of plants, E_i the experiment effect, a_i the accession effect, and β_0 an intercept. The model fitting procedure was implemented using the `glm()` procedure in R, with option `family = gaussian(link = log)`, s_i as response, t_i as offset, and β_0, E_i, a_i as fixed effects.

ACKNOWLEDGEMENTS

Funded by a EURYI-award, a SSF Future Research Leader Grant (both to Ö.C) and a FORMAS grant (to L.H.). We thank Leif Andersson, Jennifer Lachowiec and Yanjun Zan for helpful input. Also, providers of pre-publication sequence data within the 1001-genomes project are acknowledged for their efforts in creating this community resource, including Monsanto Company, the Weigel laboratory at the Max Planck Institute for Developmental Biology, the IGS of the Center for Biotechnology of the University of Bielefeld, the DOE Joint Genome Institute (JGI), the Joint BioEnergy Institute, the Nordborg laboratory of the Gregor Mendel Institute of Molecular Plant Biology, the Bergelson lab of the University of Chicago and the Ecker lab of the Salk Institute for Biological Studies, La Jolla, CA.

AUTHOR CONTRIBUTIONS

X.S. and Ö.C. conceived and designed the experiments, contributed to all analyses and wrote the paper. Ö.C. led and coordinated the study. X.S. developed the method for performing the genome-scan. S.F contributed to the analyses of the expression and methylation data. M.P. contributed to the replication association analysis. Z.S. contributed to the functional analyses of genetic

polymorphisms. L.H. and J.J. planed and conducted the heat-stress experiments. S.F., M.P., L.H., J.J. and Z.S. commented on the manuscript.

REFERENCES

1. Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, et al. (2011) Adaptation to Climate Across the Arabidopsis thaliana Genome. Science 334: 83–86.
2. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, et al. (2011) Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet 43: 956–963. doi: 10.1038/ng.911.
3. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, et al. (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Research 18: 2024–2033. doi:10.1101/gr.080200.108.
4. Schneeberger K, Hagmann J, Ossowski S, Warthmann N, Gesing S, et al. (2009) Simultaneous alignment of short reads against multiple genomes. Genome Biol 10: R98. doi:10.1186/gb-2009-10-9-r98.
5. Schneeberger K, Ossowski S, Ott F, Klein JD, Wang X, et al. (2011) Reference-guided assembly of four diverse Arabidopsis thaliana genomes. Proceedings of the National Academy of Sciences 108: 10249–10254. doi:10.1073/pnas.1107739108.
6. Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, et al. (2013) Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat Genet 45: 884–890. doi:10.1038/ng.2678.
7. Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, et al. (2013) Patterns of population epigenomic diversity. Nature 495: 193–198. doi:10.1038/nature11968.
8. Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, et al. (2011) A Map of Local Adaptation in Arabidopsis thaliana. Science 334: 86–89.
9. Pettersson ME, Nelson RM, Carlborg Ö (2012) Selection on variance-controlling genes: adaptability or stability. Evolution 66: 3945–3949. doi:10.1111/j.1558-5646.2012.01753.x.

10. Shen X, Pettersson M, Rönnegård L, Carlborg Ö (2012) Inheritance beyond plain heritability: variance-controlling genes in Arabidopsis thaliana. PLoS Genet 8: e1002839. doi:10.1371/journal.pgen.1002839.

11. Shen X, Alam M, Fikse F, Rönnegård L (2013) A novel generalized ridge regression method for quantitative genetics. Genetics 193: 1255–1268. doi:10.1534/genetics.112.146720.

12. Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, et al. (2012) Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. Nat Genet 44: 212–216. doi:10.1038/ng.1042.

13. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, et al. (2010) A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1;1. PLoS Genet 6: e1001193. doi:10.1371/journal.pgen.1001193.

14. Trontin C, Tisné S, Bach L, Loudet O (2011) What does Arabidopsis natural variation teach us (and does not teach us) about adaptation in plants? Curr Opin Plant Biol 14: 225–231. doi:10.1016/j.pbi.2011.03.024.

15. Weigel D (2012) Natural variation in Arabidopsis: from molecular genetics to ecological genomics. Plant Physiology 158: 2–22. doi:10.1104/pp.111.189845.

16. Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, et al. (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153: 193–205. doi:10.1016/j.cell.2013.02.033.

17. Stroud H, Do T, Du J, Zhong X, Feng S, et al. (2013) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21: 64–72. doi:10.1038/nsmb.2735.

18. Köhler C, Wolff P, Spillane C (2012) Epigenetic mechanisms underlying genomic imprinting in plants. Annu Rev Plant Biol 63: 331–352. doi:10.1146/annurev-arplant-042811-105514.

19. Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, et al. (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light,

drought and cold stress responses. The Plant Journal 50: 347–363. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2007.03052.x/full.

20. Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, et al. (2012) Widespread dynamic DNA methylation in response to biotic stress. Proceedings of the National Academy of Sciences 109: E2183–E2191. doi:10.1073/pnas.1209329109.

21. Geiler-Samerotte K, Bauer C, Li S, Ziv N, Gresham D, et al. (2013) The details in the distributions: why and how to study phenotypic variability. Current Opinion in Biotechnology: 1–8. doi:10.1016/j.copbio.2013.03.010.

22. Shen X, Li Y, Rönnegård L, Uden P, Carlborg Ö (2014) Application of a genomic model for high-dimensional chemometric analysis. Journal of Chemometrics: In press

23. Chen W-M, Abecasis GR (2007) Family-based association tests for genomewide association scans. Am J Hum Genet 81: 913–926. doi:10.1086/521580.

24. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R package for genome-wide association analysis. Bioinformatics 23: 1294–1296.

25. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004.

26. Shen X, Carlborg Ö (2013) Beware of risk for increased false positive rates in genome-wide association studies for phenotypic variability. Front Genet 4: 93. doi:10.3389/fgene.2013.00093.

27. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26: 2069–2070. doi:10.1093/bioinformatics/btq330.

28. Li X, Kierczak M, Shen X, Ahsan M, Carlborg Ö, et al. (2013) PASE: a novel method for functional prediction of amino acid substitutions based on physicochemical properties. Front Genet 4: 21. doi:10.3389/fgene.2013.00021.

Figures:

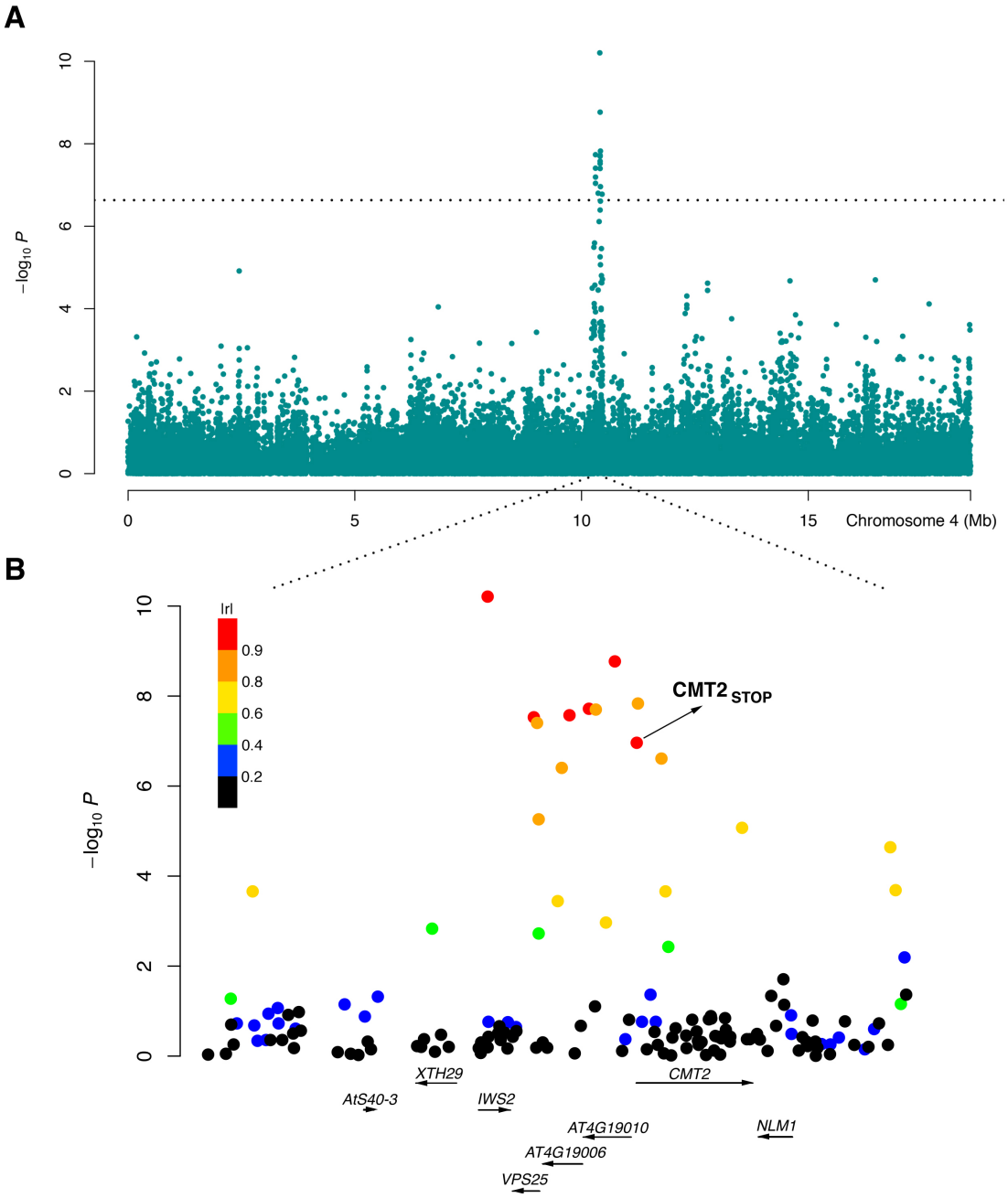


Figure 1. An LD block associated with temperature seasonality contains *CMT2*. A genome-wide significant variance-heterogeneity association signal was identified for temperature seasonality in the RegMap collection of natural *Arabidopsis thaliana* accessions [1]. The peak on chromosome 4 around 10 Mb **(A)** mapped to a haplotype block **(B)** containing a nonsense mutation (*CMT2STOP*) early in the first exon of the Chromomethylase 2 (*CMT2*) gene. Color coding based on |r| (the absolute value of the correlation coefficient) as a measure of LD between each SNP in the region and the leading SNP in the association analysis.

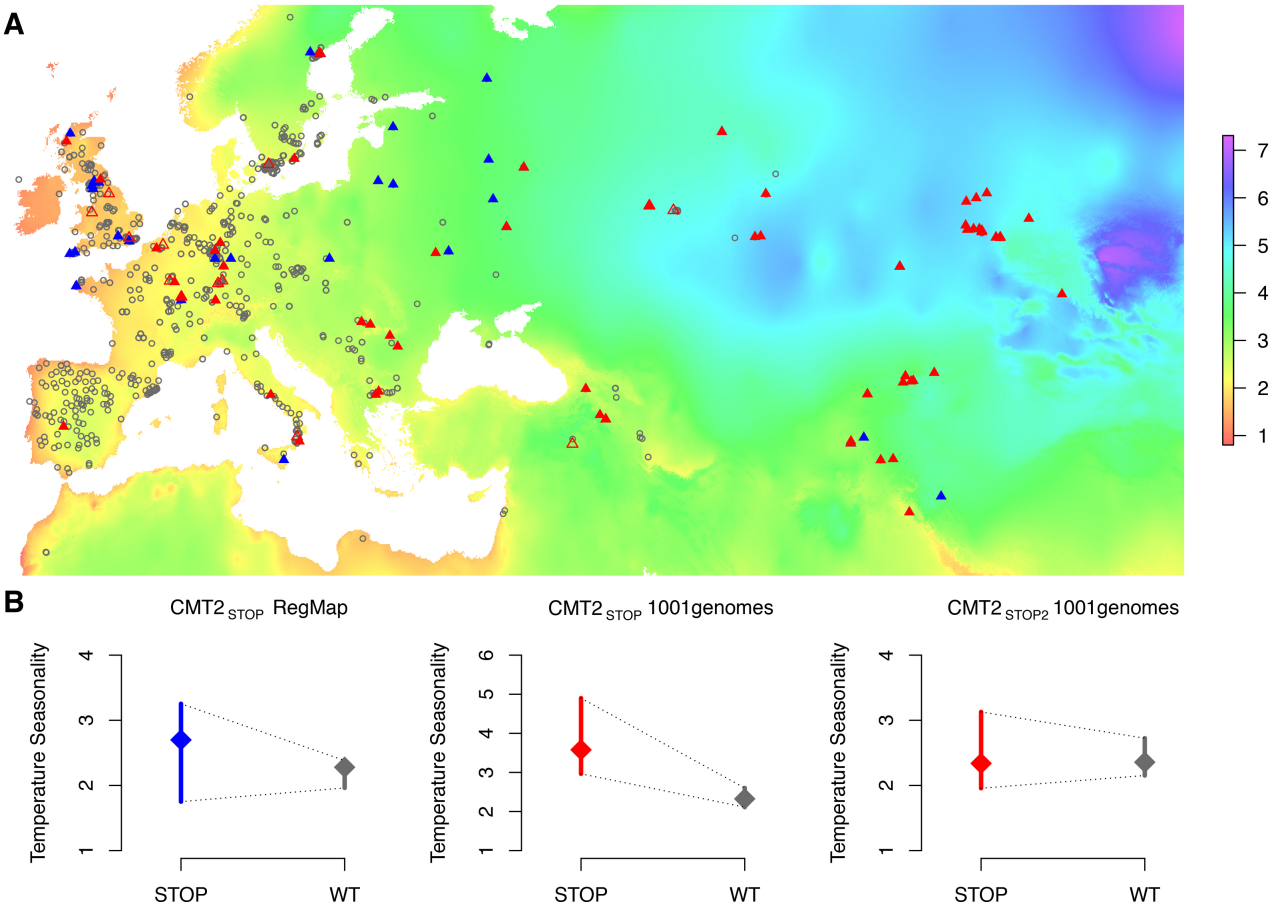


Figure 2. Geographic distribution of, and heterogenous variance for, three *CMT2* alleles in two collections of *A. thaliana* accessions. The geographic distributions **(A)** of the wild-type ($CMT2_{WT}$; gray circles) and two nonsense alleles ($CMT2_{STOP}$/$CMT2_{STOP2}$; filled/open triangles) in the *CMT2* gene that illustrates a clustering of $CMT2_{WT}$ alleles in less variable regions and a greater dispersion of the nonsense alleles across different climates both in the RegMap [1] (blue) and the 1001-genomes [2](red) *A. thaliana* collections. The resulting variance-heterogeneity in temperature seasonality between genotypes is highly significant, as illustrated by the quantile plots in **(B)** where the median is indicated by a diamond and a bar representing the 25% to 75% quantile range. The color scale indicate the level of temperature seasonality across the map. The colorkey in (**A**) represent the temperature seasonality values, given as the standard-deviation in % of the mean temperature (K).

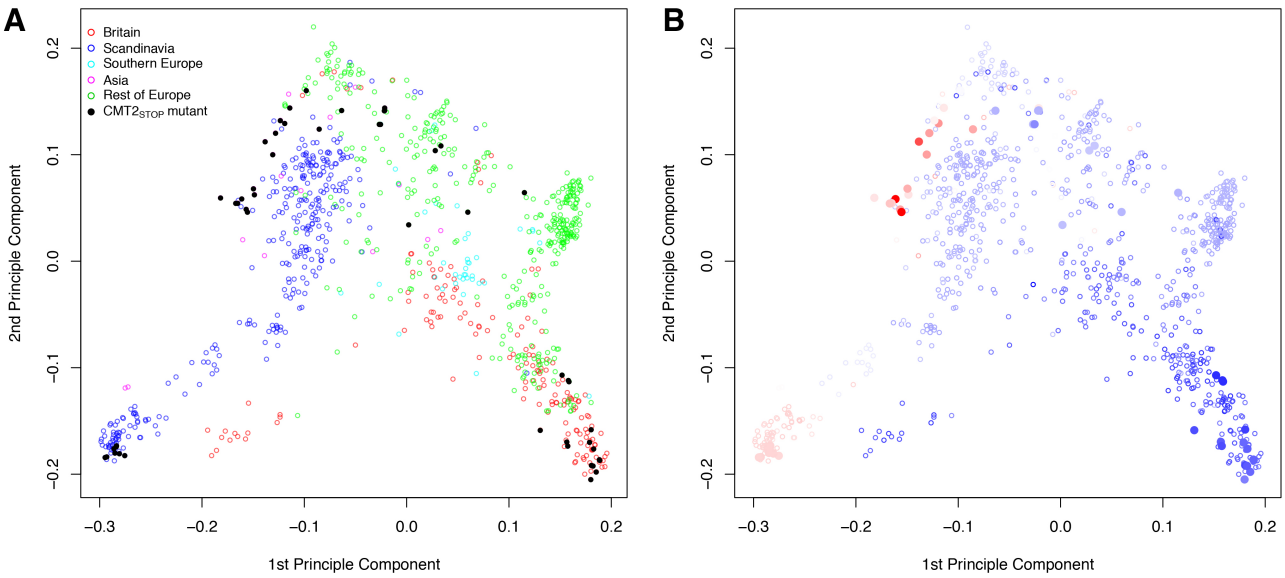


Figure 3. Principle components of the genomic kinship in the RegMap collection for the accessions carrying the alternative alleles at the Chromomethylase 2 locus (*CMT2*$_{STOP}$ and *CMT2*$_{WT}$ as filled and empty circles, respectively). Coloring is based on **(A)** geographical regions (defined as in Supplementary Figure 37) and **(B)** temperature seasonality, ranging from dark blue (least variable) to red (most variable).

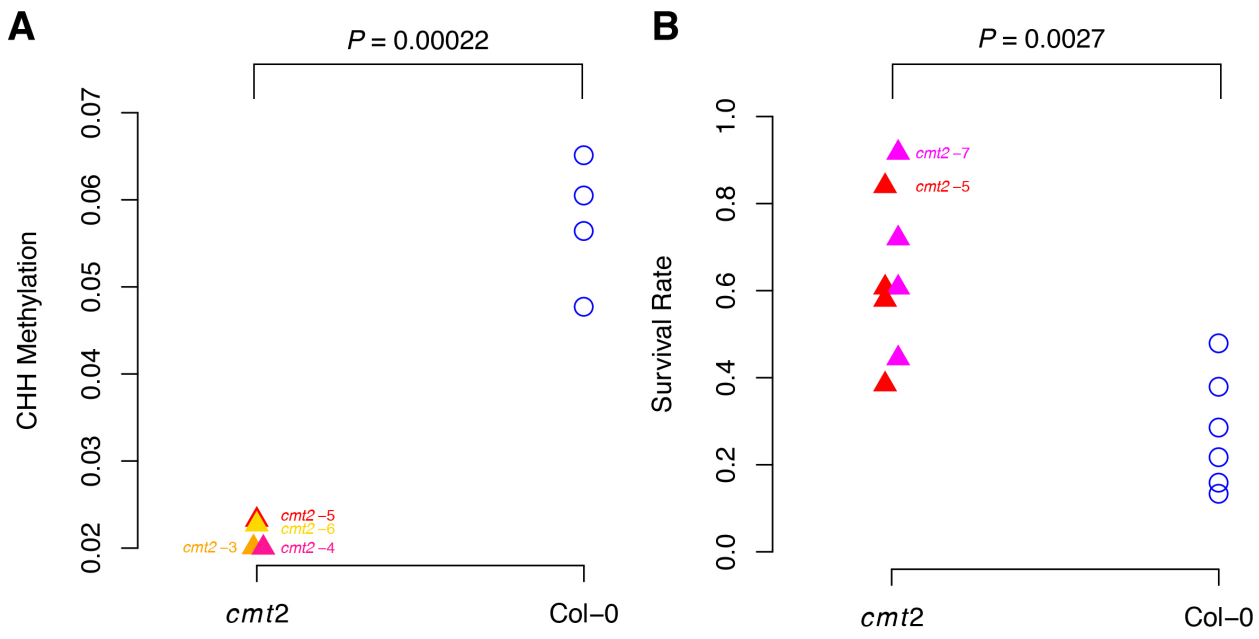


Figure 4. Differential *CMT2*-mediated CHH methylation pattern, and rate of survival under heat-stress, of *cmt2* mutants. **A.** The CHH methylation in *cmt2* (*cmt2-3, cmt2-4, cmt2-5* and *cmt2-6*) plants is significantly lower at the CHH sites mapped in the methylome-wide association analysis between the natural accessions carrying the $CMT2_{STOP}$ and $CMT2_{WT}$ alleles, implying $CMT2_{STOP}$ as a natural *cmt2* allele. **B.** The survival rate is significantly higher for *cmt2* (*cmt2-5* and *cmt2-7*) than for Col-0 plants under heat-stress (24 h at 37.5°C), illustrating the functional role of *CMT2* in temperature-stress response. P-values in A and B were obtained using a log-linear regression.

Table:

Table 1. Loci with genome-significant, non-additive effects on climate adaptation and a functional analysis of nearby genes ($r^2 > 0.8$) containing missense or nonsense mutations.

Trait	Leading SNP					Selected Candidate Genes			Mutant Analysis[1]	
	Chrom	Pos (bp)	MAF[2]	P-value[3]	Effect[4]	Gene name	Locus	# Mut[5]	PASE	MSA
Temperature seasonality										
	2[a]	12 169 701	0.08	2.0E-08	0.53±0.08	*BGAL8*	*AT2G28470*	4	0.36[6]	0.72[6]
	4	10 406 018	0.05	6.1E-11	0.51±0.06	*IWS2*	*AT4G19000*	1	0.21	0.35
						CMT2	*AT4G19020*	7	STOP	STOP
Maximum temperature in the warmest month										
	1	6 936 457	0.05	1.9E-07	0.34±0.07		*AT1G19990*	1	0.64	0.2
Minimum temperature in the coldest month										
	2[b]	18 620 697	0.08	3.5E-08	0.33±0.05					
	2[c]	19 397 389	0.05	5.0E-08	0.38±0.06					
	5	14 067 526	0.07	4.2E-08	0.33±0.05		*AT5G35930*	1	0.30	0.05
	5[d]	18 397 418	0.11	1.2E-07	0.28±0.05					
Number of consecutive cold days										
	2[b]	18 620 697	0.08	1.7E-07	0.33±0.05					
	2[c]	19 397 389	0.05	7.2E-08	0.39±0.06					
	5	7 492 277	0.08	4.3E-09	0.38±0.06		*AT5G22560*	4	0.63	0.11
	5[d]	18 397 418	0.11	1.9E-07	0.29±0.05					
Day length in spring										
	2[a]	12 169 701	0.08	2.0E-07	0.46±0.08	*BGAL8*	*AT2G28470*	4	0.36	0.72
	3	12 642 006	0.07	9.4E-08	0.29±0.05					
	4[e]	14 788 320	0.08	2.2E-08	0.39±0.06	*VEL1*	*AT4G30200*	2	0.26	0.06
Relative humidity in spring										
	3	1 816 353	0.07	1.2E-08	0.39±0.06					
	4[e]	14 834 441	0.06	6.4E-08	0.49±0.08	*VEL1*	*AT4G30200*	2	0.26	0.06
					0.43±0.07	*XTH19*	*AT4G30290*	1	0.14	0.49
	5	8 380 640	0.07	6.3E-08						
Length of the growing season										
	3	576 148	0.08	1.4E-07	0.27±0.04					
Number of consecutive frost-free days										
	1	953 031	0.24	8.0E-08	0.25±0.04	*SOM*	*AT1G03790*	3	0.25	0.55
	1	6 463 065	0.08	1.9E-07	0.33±0.06					
	2	9 904 076	0.22	2.2E-08	0.22±0.04					

[a,b,c,d,e]Loci affecting affect multiple traits; [1]The predicted functional effect score for the strongest mis-sense mutation in the gene based on amino-acid physiochemical properties (PASE) and evolutionary conservation (MSA) [28]; [2]MAF: Minor Allele Frequency; [3]P-value: significance after genomic-control from a linear regression analysis of squared z-scores accounting for population stratification. [4]Effect: Standardized genetic effect on adaptability (Chi-square distributed) ± standard error (unit: phenotypic standard deviations); [5]#Mut: number mis- and non-sense mutations in the gene in the 1001-genomes dataset [2]. [6]Locus contains two missense mutations with equally strong predicted effects

Supplementary Information for

Natural *CMT2* variation is associated with genome-wide methylation changes and temperature adaption

Xia Shen, Jennifer De Jonge, Simon Forsberg, Mats Pettersson, Zheya Sheng,
Lars Hennig and Örjan Carlborg*

*To whom correspondence should be addressed. E-mail: orjan.carlborg@slu.se

This PDF file includes:

Supplementary Text
Supplementary Fig. 1-44
Supplementary Table 1-2

Other Results

CMT2 is a potential target for the nonsense-mediated RNA decay (NMD) pathway

To further explore the potential mechanism underlying the observed effect of *CMT2*, as well as the heterogeneity within the group of mutant accessions, we also studied the level of mRNA in the plants. The motivation for this evaluation was that the $CMT2_{STOP}$ allele will produce an mRNA with a premature translation termination codon, which makes it a likely target for the nonsense-mediated mRNA decay (NMD) pathway. Thus, the expectation is that accessions carrying this allele would have lower transcript abundance than the wild-type. For this study, we used data from two studies that contained data both on the genotype for *CMT2* as well as RNA-seq data for the same lines. First, we analyzed the 19 genomes project data[28] that contained full genome sequences and transcriptomes for 19 *A. thaliana* accessions, 2 of which (Ct-1 and Kn-0) are part of the RegMap panel and carries the $CMT2_{STOP}$ mutation according to their 250k SNP-chip genotypes[13]. Utilizing data from both the biological replicates of seedling mRNA, the difference in mRNA abundance was highly significant between mutant and wild type accessions ($P = 6.9 \times 10^{-5}$), with a higher expression in the wild-type. A similar analysis was done using a larger data set from [7], with *CMT2* genotypes obtained from SNPs called from whole-genome re-sequencing data and RNA-seq data was obtained from analyses of leaf tissue in 14 $CMT2_{STOP}$ and 92 $CMT2_{WT}$ accessions. Here, the average mRNA abundance was higher for the wild-type accessions, but the difference was not significant in the complete dataset ($P = 0.14$). However, the mRNA levels were significantly higher among the four mutant accessions that displayed a methylation pattern resembling that of the wild-type in the analysis above (t-test; $P = 0.01$) and when those lines were removed from the comparison, the levels of mRNA was significantly higher in the wild-type accessions than in the ten remaining mutants (t-test; $P = 0.02$). These results indicate that *CMT2* mRNA levels are influenced by the genotype and that it is connected to the methylation state in the plant, but provide no conclusive evidence on the functional connection between the two.

VEL1 and adaptation to day length

A thaliana is a facultative photoperiodic flowering plant and hence non-inductive photoperiods will delay, but not abolish, flowering. The genetic control of this phenotypic plasticity is thus an adaptive trait. A significant association was detected near two genes, *VEL1* and *XTH19*, containing two and one non-synonymous amino acid substitutions, respectively (Table 1). The major allele was dominant in short-day

regions, whereas the alternative allele was more plastic in relation to day-length. *XTH19* has been implied as a regulator of shade avoidance[29], but information about its potential involvement in regulation of photoperiodic length is lacking. *VEL1*, regulates the epigenetic silencing of genes in the *FLC*-pathway in response to vernalization[30] and photoperiod length[31] resulting in an acceleration of flowering under non-inductive photoperiods. A feasible explanation for the existence of an adaptability *VEL1*-allele could thus be that accelerated flowering is beneficial under short-day conditions, but that also lack of accelerated flowering is allowed. In long-daytime areas, however, accelerated flowering might be detrimental as day-length follows a latitudinal cline, where early flowering might be detrimental in northern areas where accelerated flowering when the day-length is short could lead to excessive exposure to cold temperatures in the early spring and hence a lower fitness.

Source Codes

Example R source code for calculating HEM genomic kinship matrix

Here, we use the example data in the **bigRR** package: http://cran.r-project.org/web/packages/bigRR/ to illustrate how an ordinary identity-by-state (IBS) kinship matrix can be update to a HEM genomic kinship matrix. The full theoretical details on this procedure are provided in [11].

```
# load the bigRR package
require(bigRR)

# load the example data
data(Arabidopsis)
X <- matrix(1, length(y), 1)
Z <- scale(Z)

# fitting SNP-BLUP, i.e. a ridge regression on all the markers across the genome
SNP.BLUP <- bigRR(y = y, X = X, Z = Z, family = binomial(link = 'logit'))

# calculate HEM (heteroscedastic effects model) genomic kinship matrix
w <- as.numeric(SNP.BLUP$u^2/(1 - SNP.BLUP$leverage))
wZt <- sqrt(w)*t(Z)
```

```
G <- crossprod(wZt)
```

Temperature seasonality phenotype preparation in R

We present the source code for phenotyping of temperature seasonality in Euro-Asia. The data downloaded from http://www.worldclim.org/ were processed using the following code to obtain an object readable by the **raster** package: http://cran.r-project.org/web/packages/raster/.

```
# load the original data files
bil_files <- grep(".bil", dir("tmean_30s_bil Folder/"), value = T)
bil_file_order <- as.numeric(sub(pattern = ".+_([0-9]+).bil",
                  x = bil_files, replacement = "\\1"))
bil_files <- bil_files[order(bil_file_order)]

# create rasters
WorldClim_stack <- stack()
for (bil_file in bil_files){
r <- raster(paste("tmean_30s_bil Folder/",bil_file, sep = ""))
WorldClim_stack <- addLayer(WorldClim_stack, r)
}

# temperature seasonality calculation
r_mean <- calc(WorldClim_stack, mean)
save(r_mean, file = "WorldClim_mean.Rdata")
writeRaster(r_mean, file = "WorldClim_mean.raster")
r_sd <- calc(WorldClim_stack, sd)
save(r_sd, file = "WorldClim_sd.Rdata")
writeRaster(r_sd, filename = "WorldClim_sd.raster")
r_mean_corr <- r_mean/10 + 273.15
save(r_mean_corr, file = "WorldClim_mean_corr.Rdata")
r_coeff_var <- 100*(r_sd/10)/r_mean_corr

# output to a raster object
```

```
writeRaster(r_coeff_var, file = "WorldClim_coeff_var.raster")

# phenotyping at given coordinates
# (LONGITUDE and LATITUDE already loaded)
require(raster)
world_temp_seas <- raster('WorldClim_coeff_var')
temp_seas <- raster::extract(world_bio5, cbind(LONGITUDE, LATITUDE))
```

Further General Discussions

Statistical properties of the vGWAS in relation to population stratification

Inherent properties of the variance heterogeneity test decreases risk of identifying locally adapted alleles An important property of the variance heterogeneity GWAS analysis is that it is inherently more powerful in detecting loci where the minor allele is associated with a higher variance than the major allele. In practice, there is no power to detect low-variance minor alleles in a GWAS setting[10,22]. Hence, the method facilitates detection of alternative (*i.e.* minor) alleles associated with a broader range of the climate variables than the reference (*i.e.* major) alleles. The method is powerful in finding associations to minor alleles associated with a broader range of climate-variables than the reference. Such alleles will, by definition, be present across a large proportion of the global population and due to this be considerably less affected by population structure. In **Figure 3** or **Supplementary Figure 15** (top panel), we illustrate this property for the inferred loci using the *CMT2* locus as an example. The MDS-plot visualizes the distribution of the $CMT2_{STOP}$ allele across the population structure present in the RegMap collection using the pairwise genome-wide relationship between the accessions based on the first two principle-components of the kinship matrix. The link between the geographic origin of the accessions and kinship is visualized by coloring the dots for each accession based on geographic origin. As expected, accessions from nearby regions (*e.g.* UK, Scandinavia and mainland Europe) are more related. The $CMT2_{STOP}$ allele is, however, not heavily confounded with population-structure and is present in most major sub-groups of the population (albeit with a higher frequency in Asia - see **Supplementary Fig. 36**).

Mixed models based vGWAS analyses to account for population structure via modeling of genome-wide kinship We statistically deal with the strong correlations that exist between climate & population

structure (*e.g.* along east-west/north-south clines) using a mixed-model based approach accounting for genomic kinship combined with genomic control. This approach has earlier been shown to control type I errors (*i.e.* genome-wide *P*-value inflation) in structured populations. The major challenge in analyzing this population is thus not the false-positive rate, as also standard GWAS analyses can be implemented in the same mixed-models framework, but rather to avoid unacceptably high type II errors (*i.e.* low power) for traits confounded with population-structure. Traditional GWAS analyses model alleles to have a linear relationship with climate, which in practice means that they mostly coincide with the population-structure along geographic clines. Hence, analyses will either be prone to identify false associations (when population-structure is not accounted for), or be under-powered (when accounting for population-structure). Although this is not explicitly discussed in the earlier reports based on this data, this is the primary reason for their lack of genome-wide significant associations to individual adaptive loci. As illustrated in **Figure 3**, the variance-heterogeneity test identifies loci present across population strata, where the signal therefore remain even after accounting for population structure via the mixed-model approach. The independence between the effect of the inferred locus and population structure can be evaluated statistically by fitting a linear mixed model where the genotype is regressed on the genomic kinship, where the heritability differs from 0 when confounding is present. For *CMT2* this estimate is zero, showing that the *CMT2* genotype is not confounded with population structure in this data.

On the power of vGWAS and GWAS analyses in highly structured populations There are several reasons for why a low overlap is expected between the results from traditional GWAS/selective-sweep analyses (as performed earlier) and the variance-heterogeneity GWAS (vGWAS) used here. First, in the absence of population stratification, the GWAS is more powerful than the vGWAS. In the presence of population stratification, however, loci affecting the mean phenotype will often be highly confounded with population-structure as they are a main genetic mechanism leading to local adaptation. In order to infer such loci when controlling for population structure, the same alleles need to have been under selection in multiple, unrelated populations, which is apparently a rare event as no such loci could be detected in the earlier studies of climate adaptation. The population genetics forces acting on variance-controlling loci are still poorly explored. Studies have, however, shown that high- and low-variance alleles are likely to co-exist in the population over extended periods of time at a frequency balanced depending on fluctuations in the surrounding environment that the population adapts to[9]. Due to this, both alleles are more likely to be present across different population strata than mean affecting alleles and therefore be less confounded with population structure. In **Supplementary Figure 36**, we exemplify

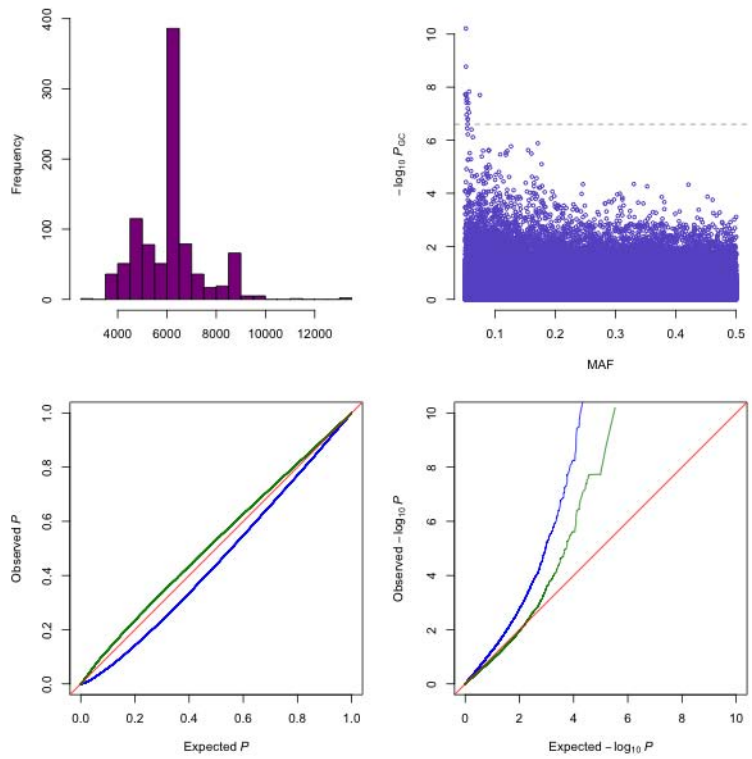
this by visualizing the allele-frequency of the $CMT2_{STOP}$ allele across the major sub-populations in the RegMap and 1001-genomes data. Although the uneven sampling across the regions in the 1001-genomes makes allele-frequency estimates uncertain, the overall picture shows that minor allele is present across all sub-populations at a lower frequency and that it has increased in frequency in Asia.

Second, a traditional GWAS searches for difference in means between genotypes, whereas the vGWAS searches for differences in variances between genotypes. As these are two different statistical properties of the phenotypic distribution, the basic assumption is that they are both statistically and biologically unrelated and consequently the loci identified by the two methods are not expected to overlap. Although some degree of overlap might be expected, *e.g.* in situations where the variance scales with the mean, the high-significance required to reach genome-wide significance in the testing, in practice only loci with strong, pure effects on one of the statistical moments seem to be able to reach such significance levels (see e.g.[10]). Formal comparisons between the results in the association studies will thus be misleading to the readers, as these will indicate that an overlap is to be expected. Results from evaluations of the overlap for sub-GWAS signals to explore the potential overlap of loci with weaker effects on both the mean and the variance shows some overlap (**Supplementary Fig. 40-44**). It should be noted, however, that comparisons of overlap at individual loci is not appropriate at these significance levels due to the lack of proper control of the type I error rate. The overall conclusions from these comparisons is i) that the power is generally very low for the GWAS after control for population stratification and ii) that even the sub-GWAS overlap is low for the two methods, but the overlap that exists is consistent with the correlation between the climate variables.

Third, the earlier studies have also inferred loci using traditional selective-sweep mapping. These analyses are designed to infer hard selective-sweeps where (potentially) adaptive alleles are assumed to have increased in frequency due to directional selection. As discussed above, the population genomic dynamics of plastic alleles does not follow the same pattern as for alleles affecting the mean (see e.g.[9]), leading to a co-existence of the alleles over prolonged periods of time. This means that they will not be surrounded by a traditional genomic footprint of directional selection that can be detected in a selective-sweep analysis and one would not expect any overlap between the loci inferred in the selective sweep and vGWAS based analyses.

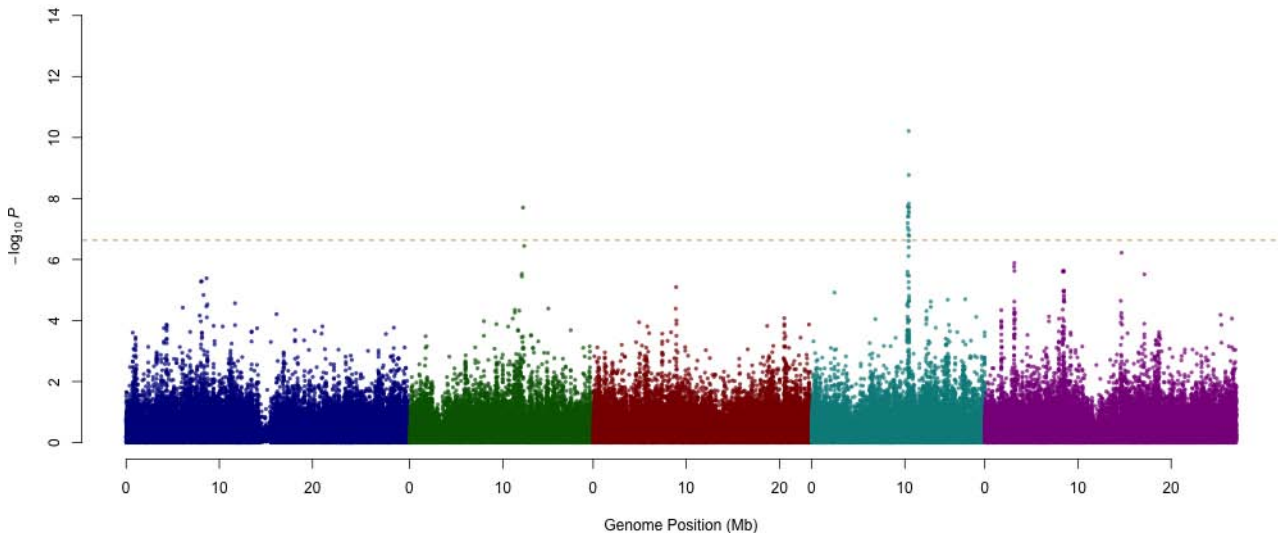
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

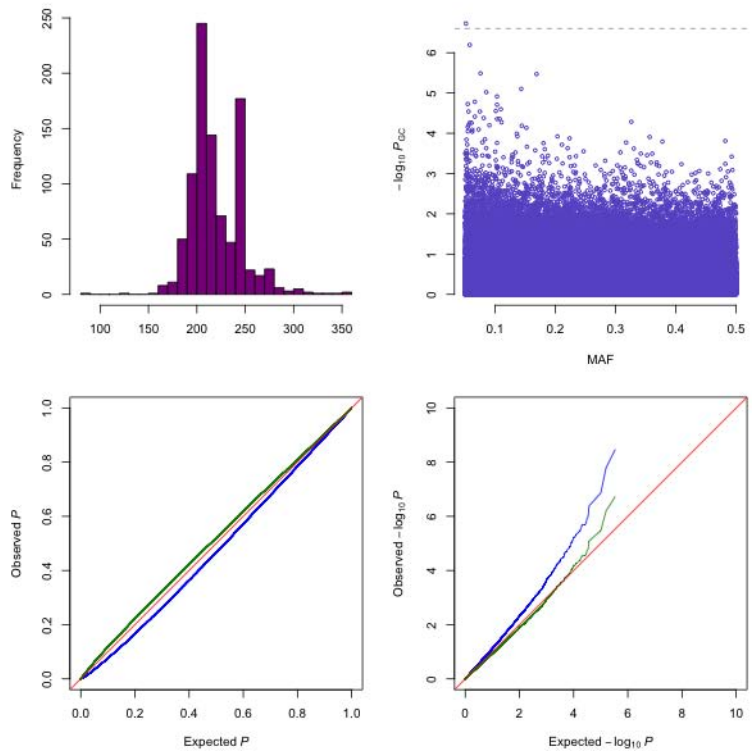
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 1 - Summary of results for temperature seasonality.

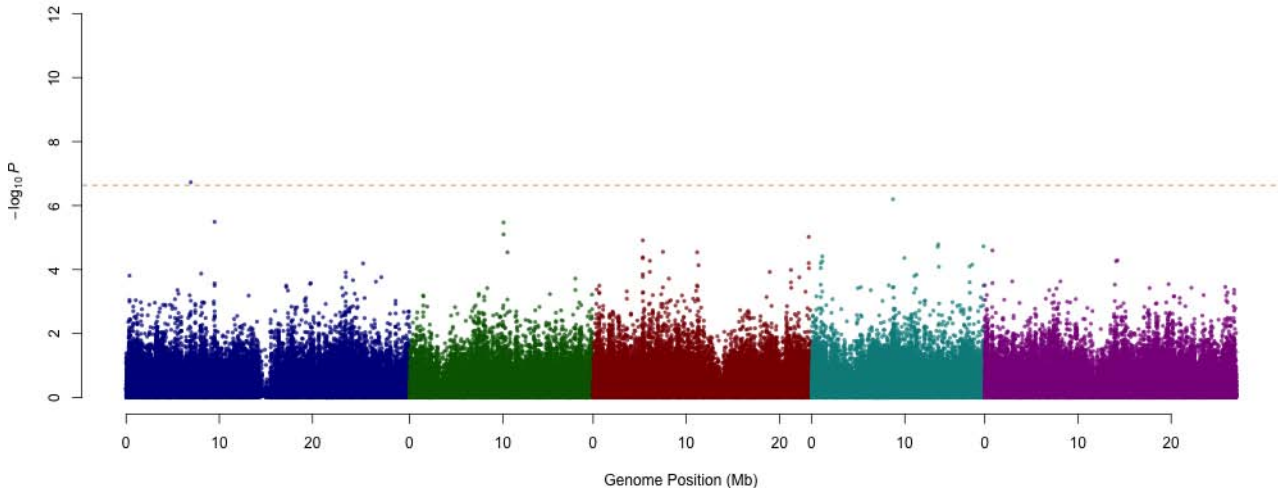
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

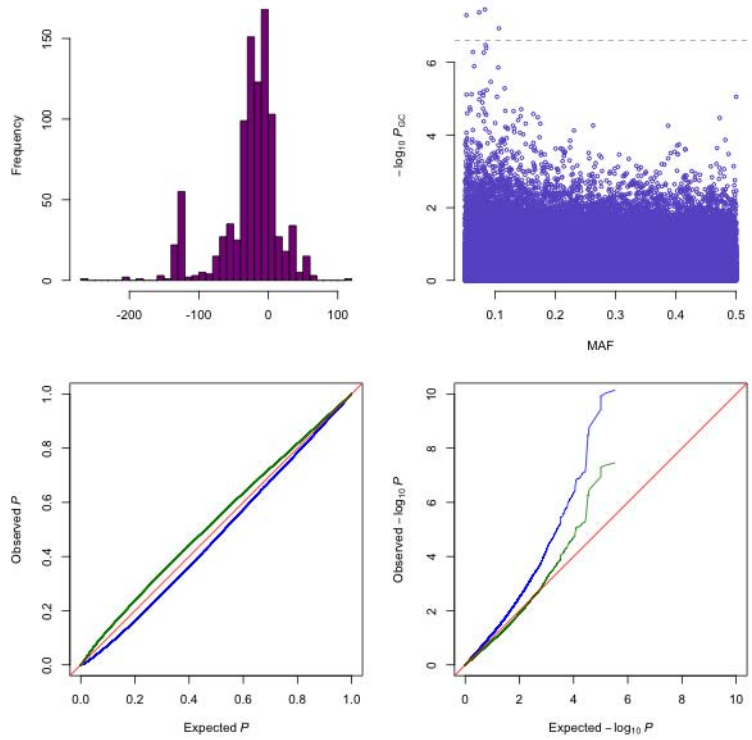
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 2 - Summary of results for maximum temperature in the warmest month.

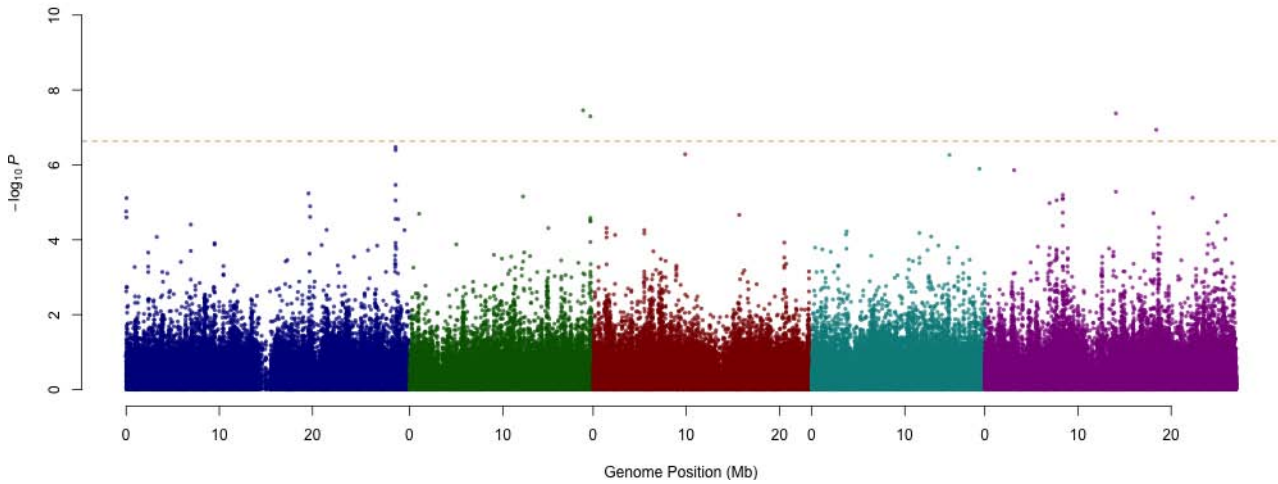
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

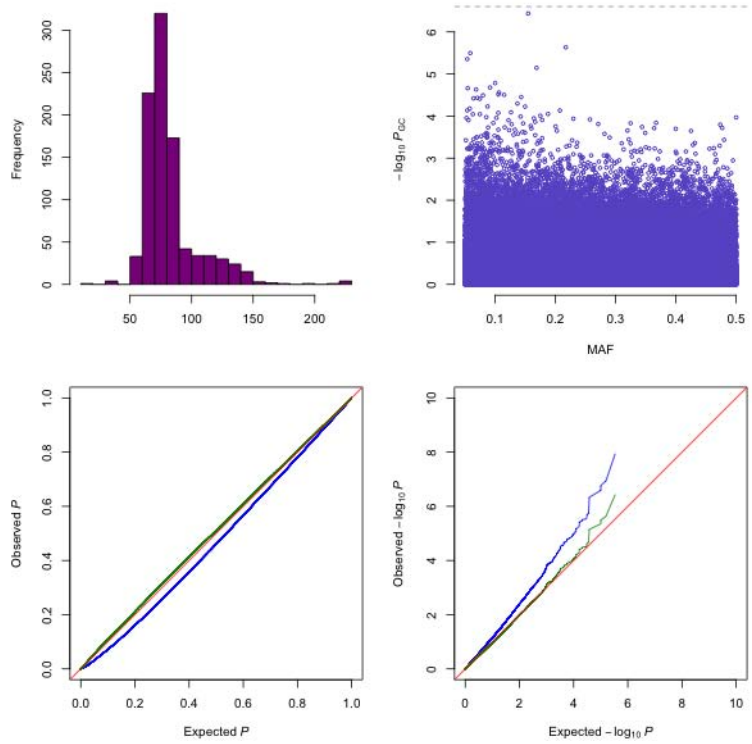
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 3 - Summary of results for minimum temperature in the coldest month.

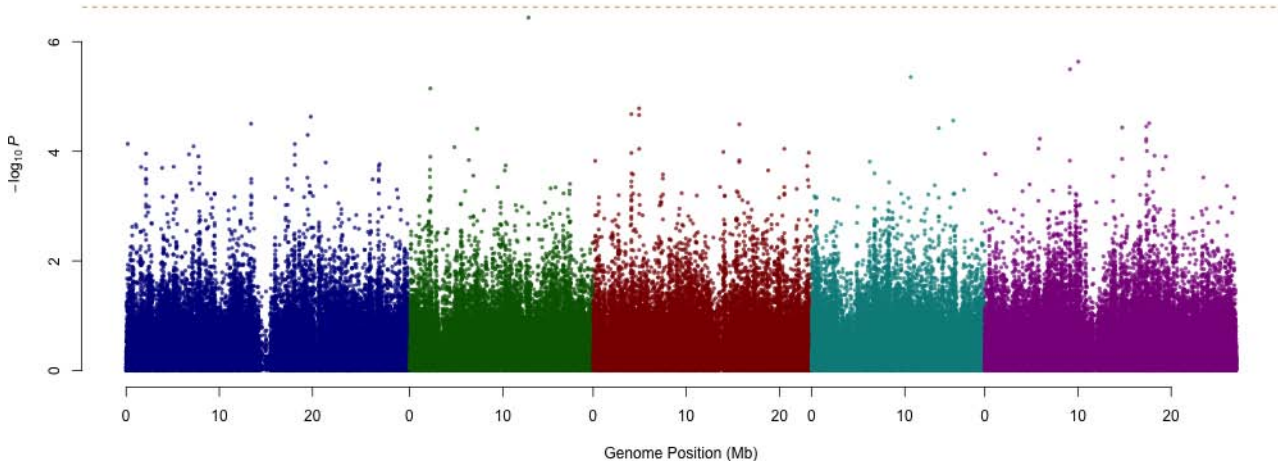
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

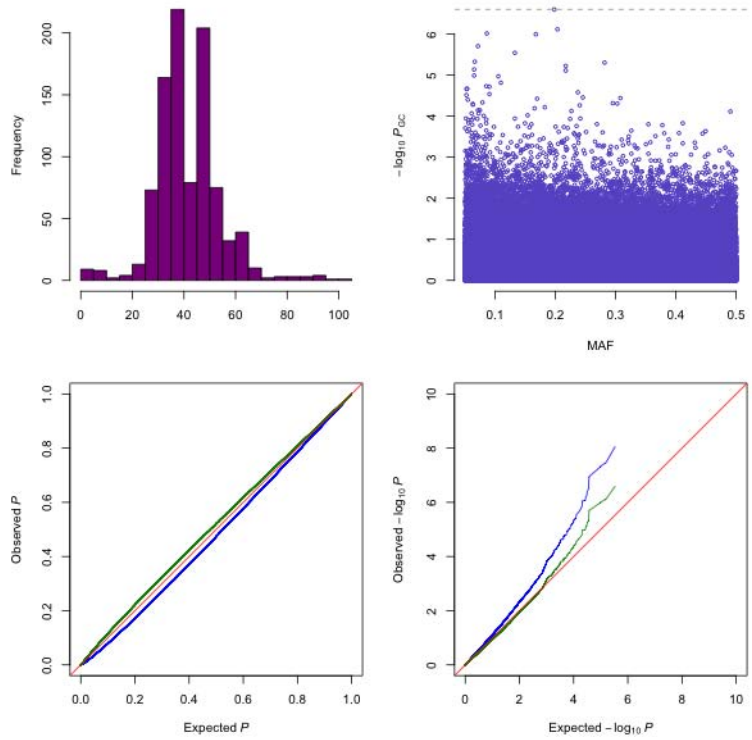
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 4 - Summary of results for precipitation in the wettest month.

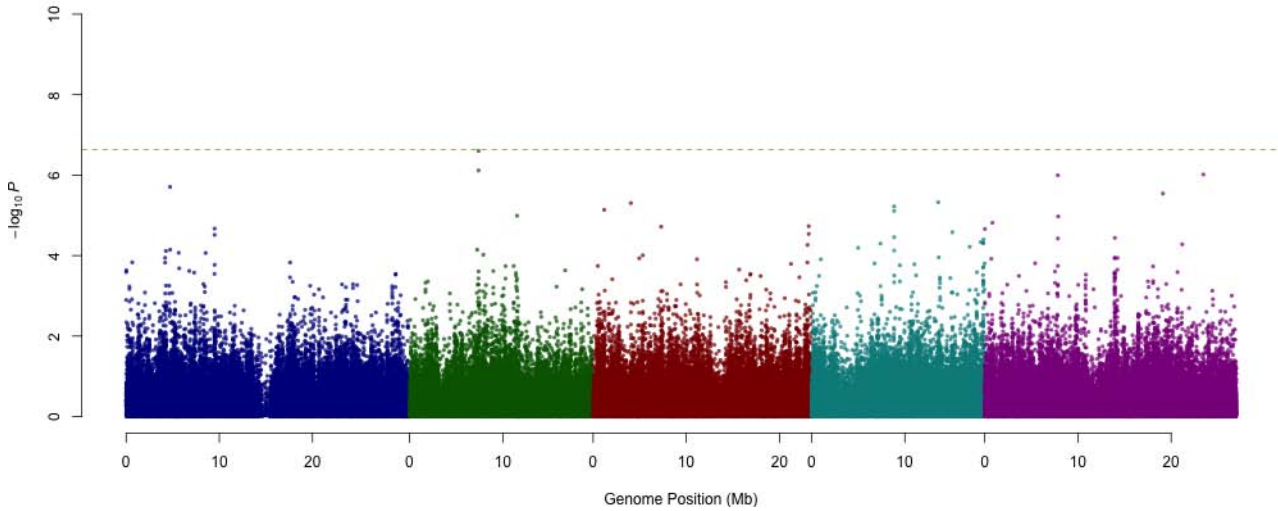
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

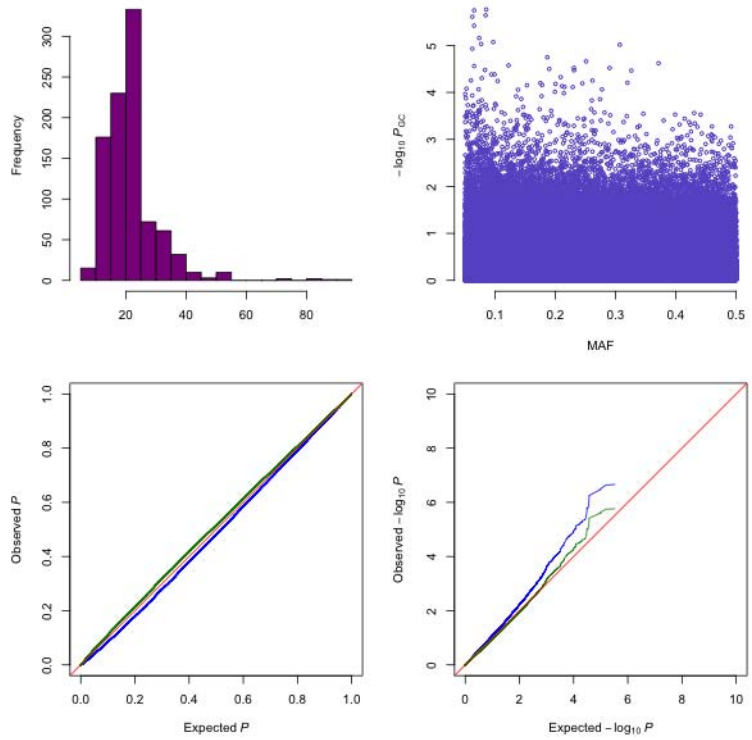
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 5 - Summary of results for precipitation in the driest month.

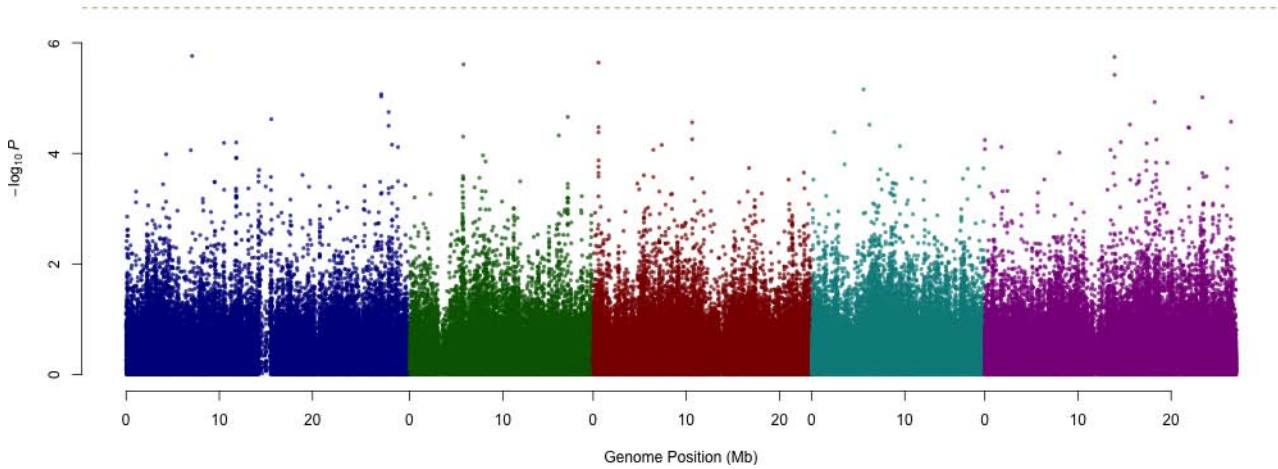
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

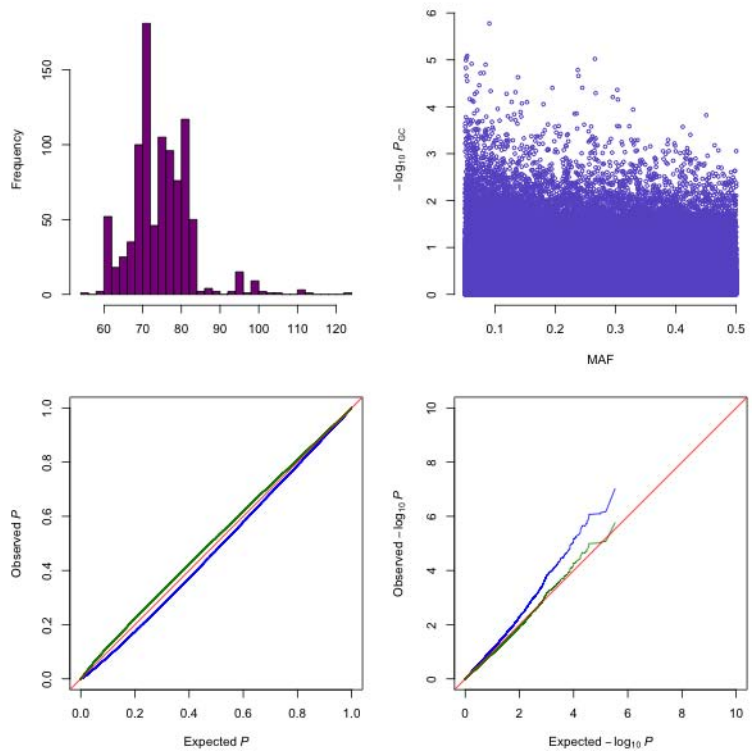
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 6 - Summary of results for precipitation CV.

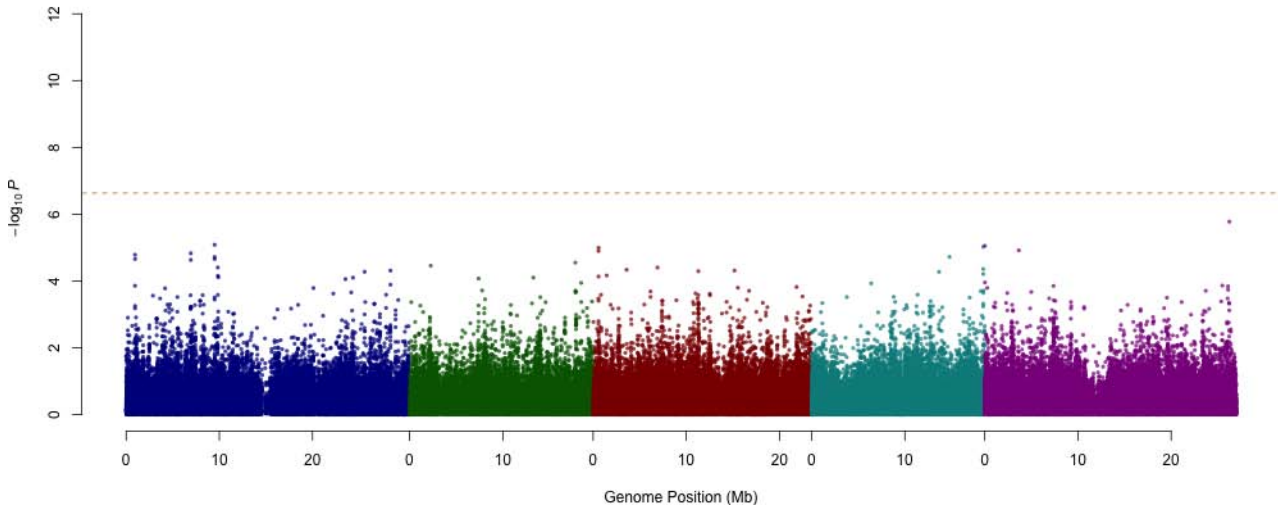
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

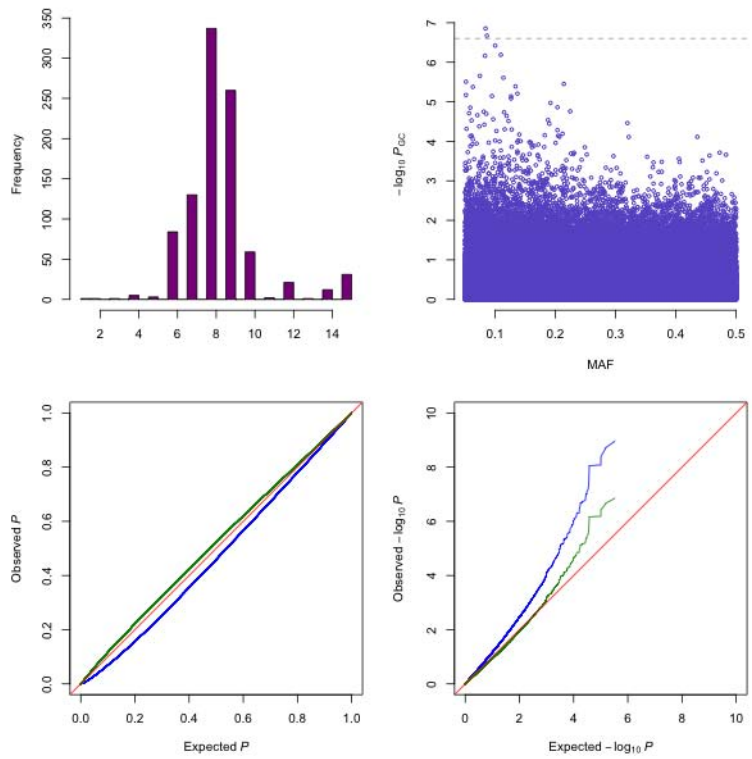
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 7 - Summary of results for photosynthetically active radiation in spring.

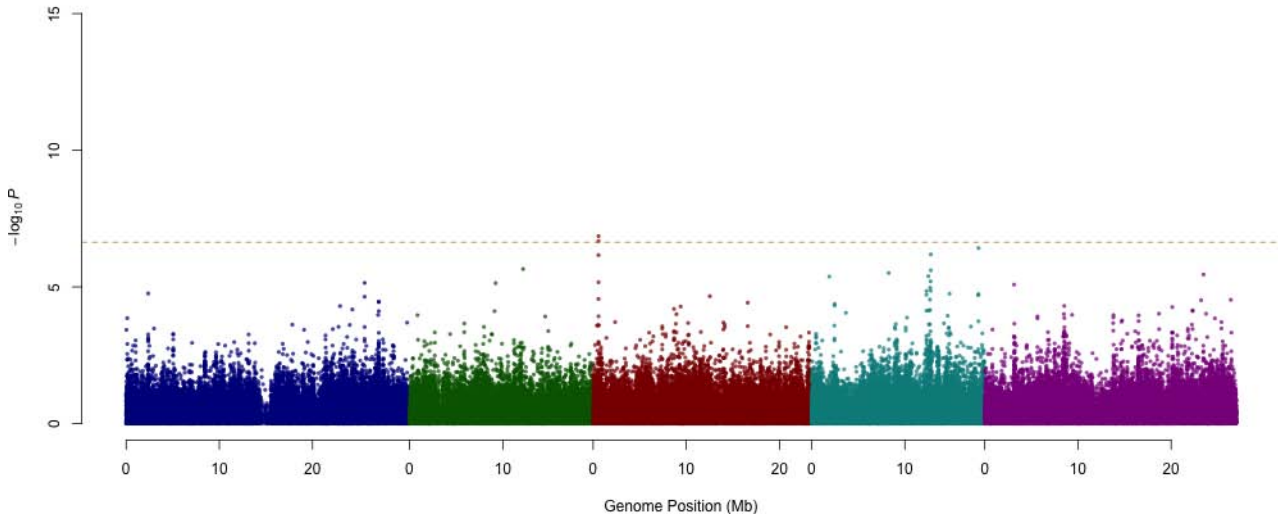
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

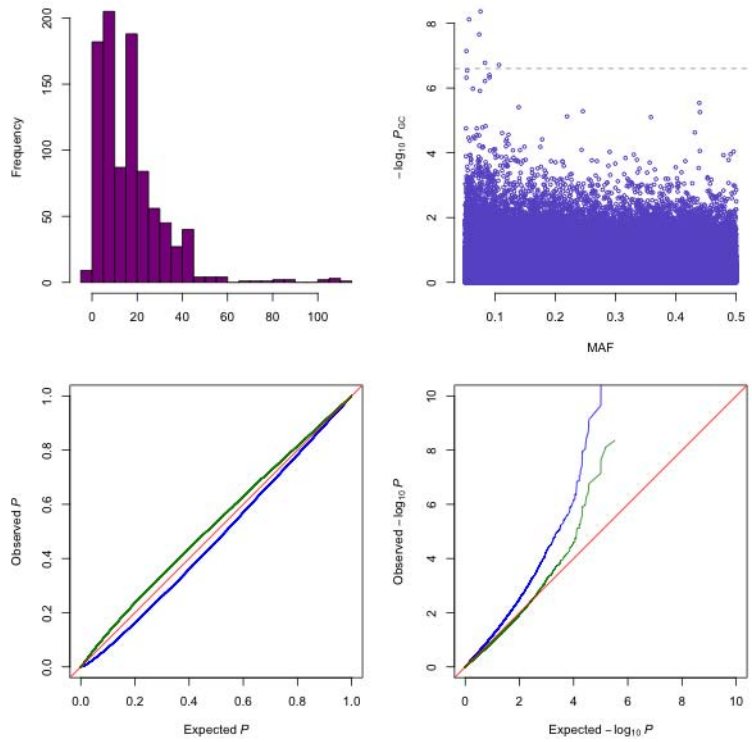
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 8 - Summary of results for length of the growing season.

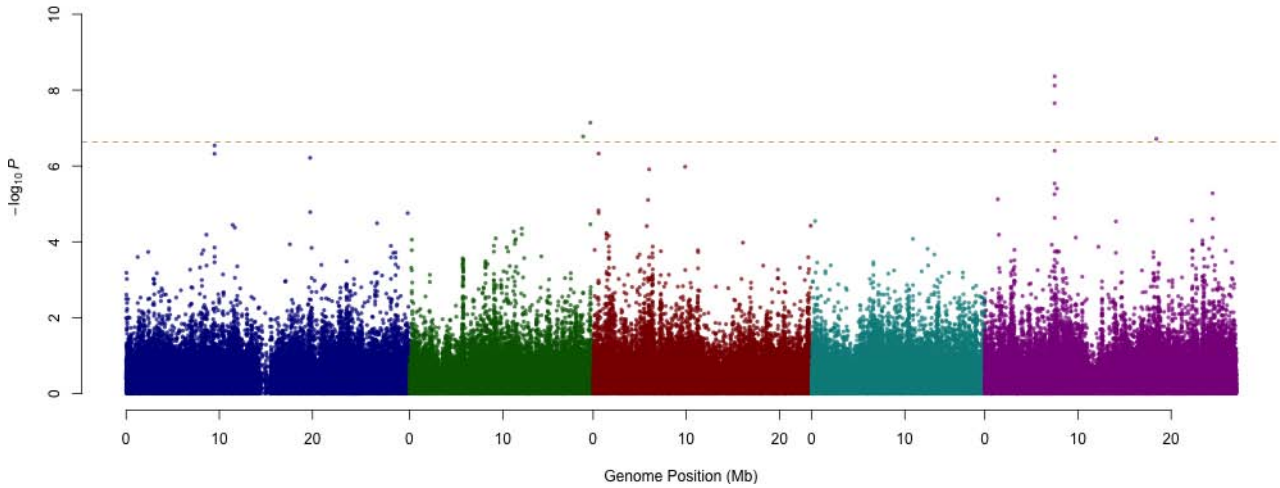
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

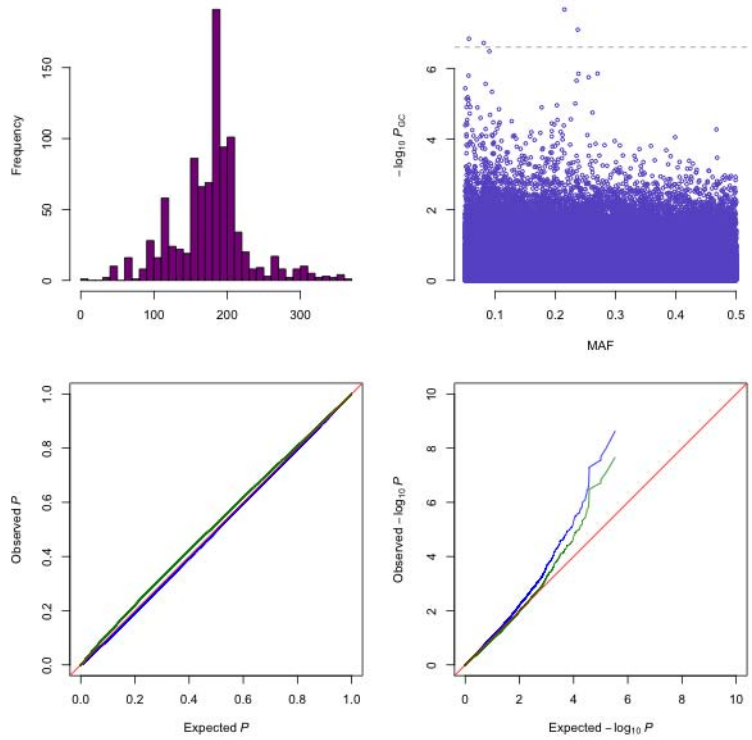
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 9 - Summary of results for number of consecutive cold days.

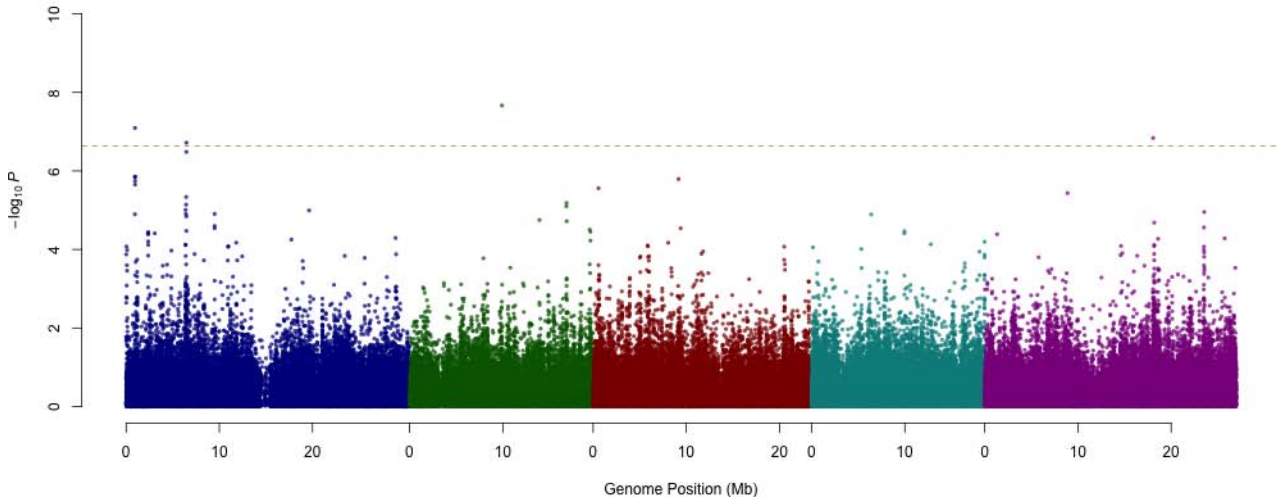
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

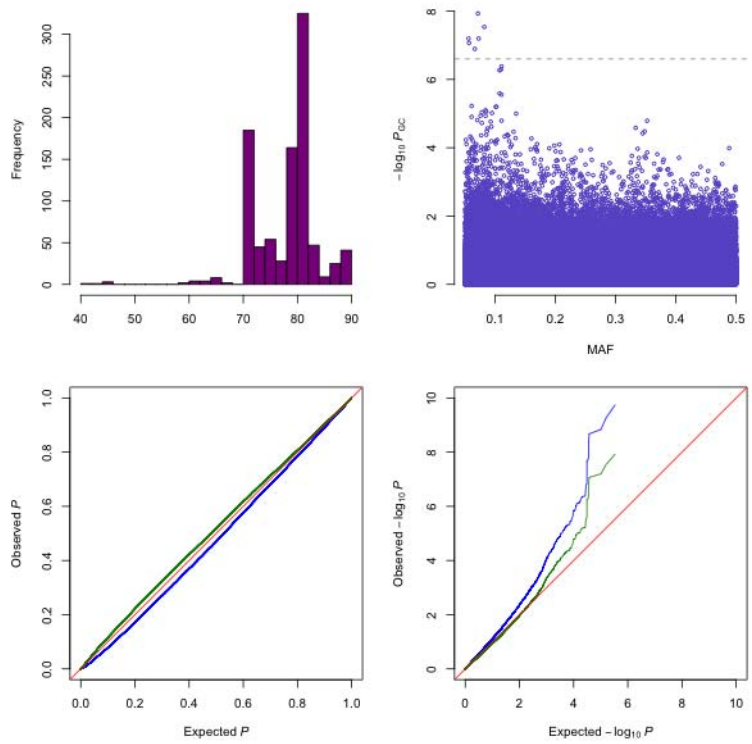
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 10 - Summary of results for number of consecutive frost-free days.

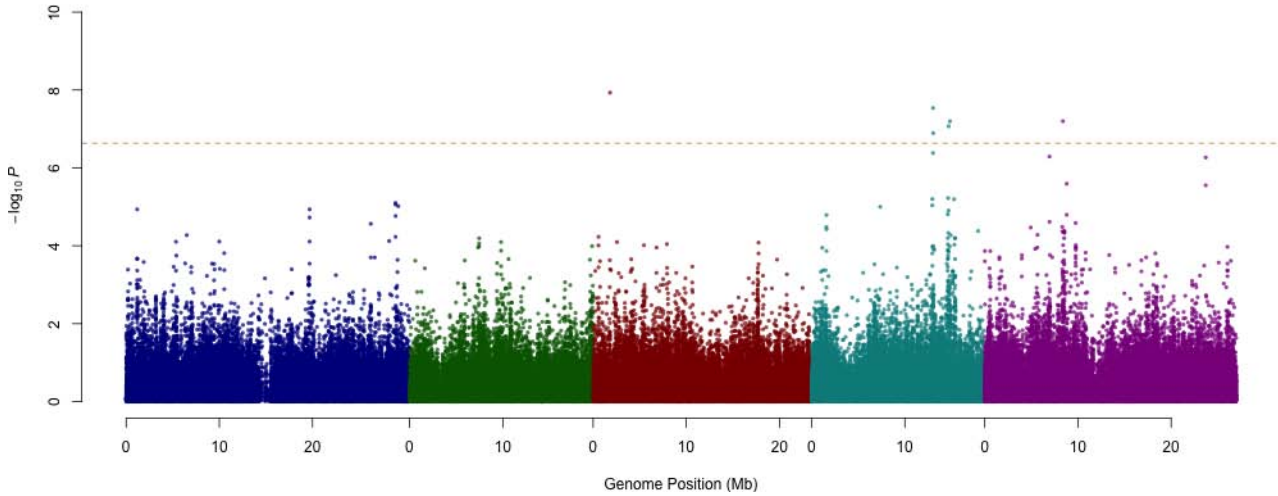
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

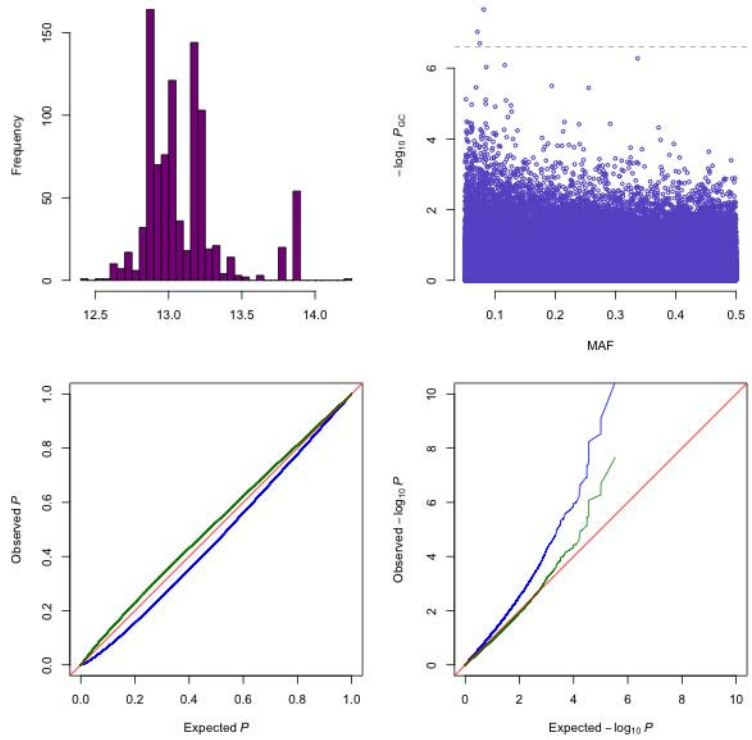
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 11 - Summary of results for relative humidity in spring.

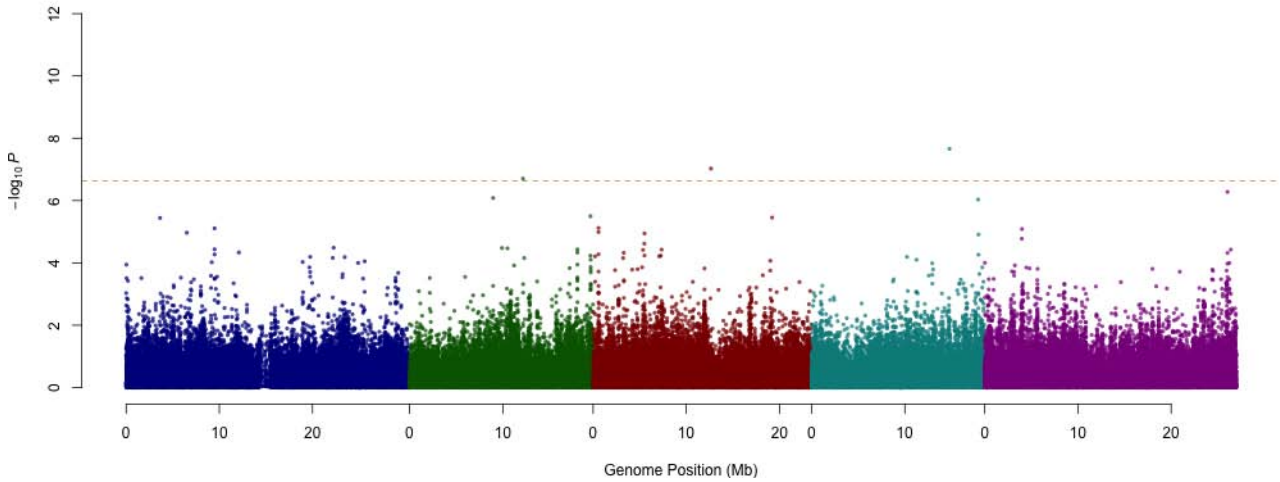
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



b: Genome-wide association mapping for climate adaptability.

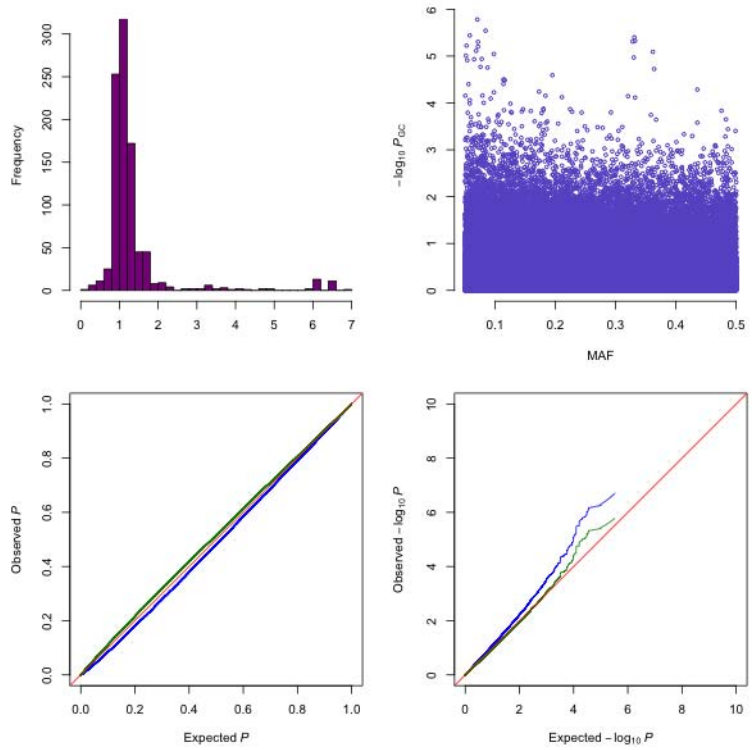
The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



Supplementary Figure 12 - Summary of results for daylength in spring.

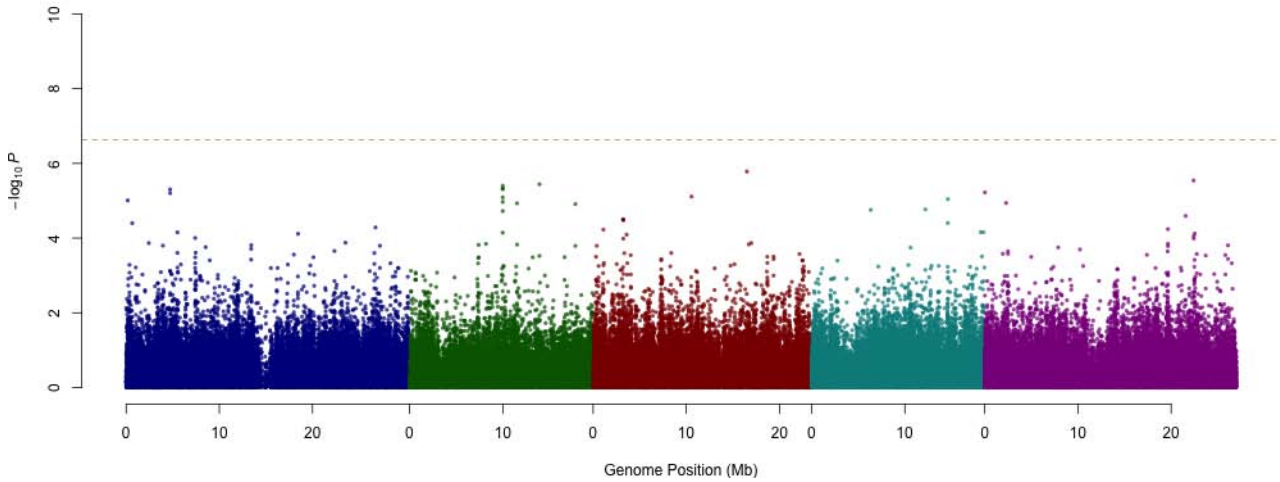
a: Phenotypic and *p*-value distributions.

Top-left: phenotypic distribution; Top-right: -$\log_{10}p$-values after genomic control (GC) against minor allele frequencies (MAF); Bottom panels: Quantile-quantile plots of *p*-values and -$\log_{10}p$-values before (blue) and after (green) GC.



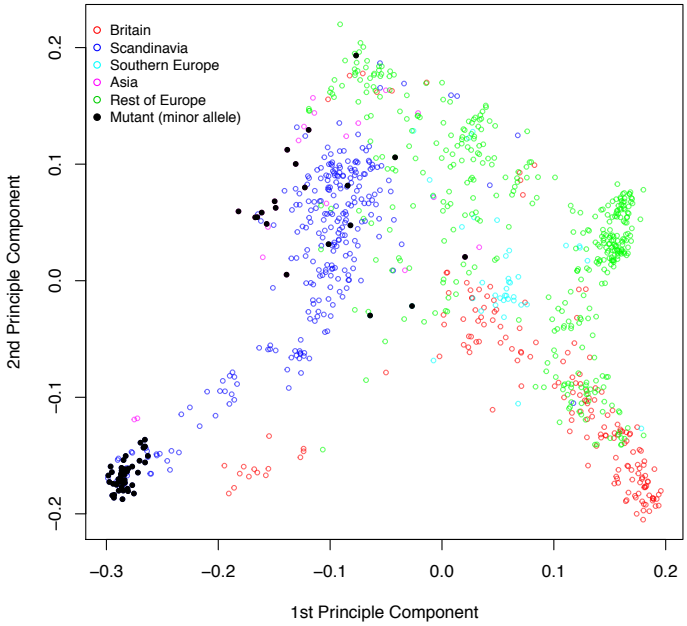
b: Genome-wide association mapping for climate adaptability.

The plotted -$\log_{10}p$-values are genomic controlled. Markers with minor allele frequencies less than 5% are removed. Chromosomes are distinguished by colors. The Bonferroni-corrected significance threshold is marked by the horizontal line.



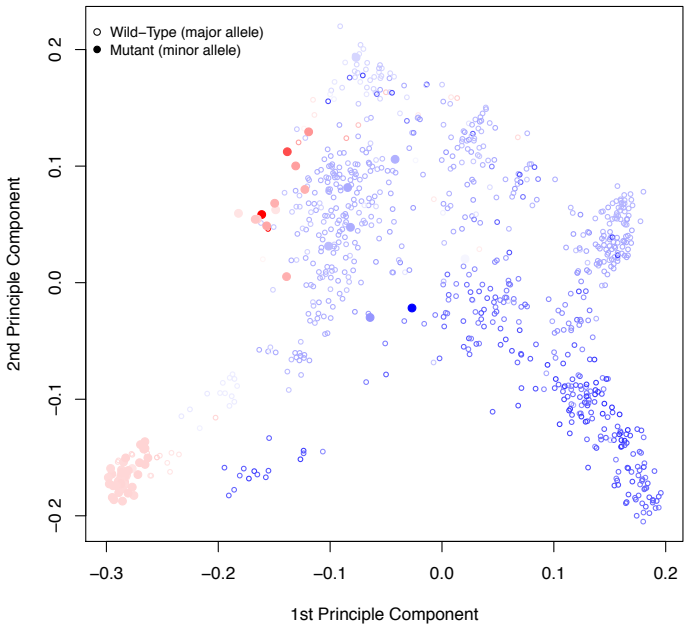
Supplementary Figure 13 - Summary of results for aridity index.

Genomic kinship principle components categorized based on geographical regions.



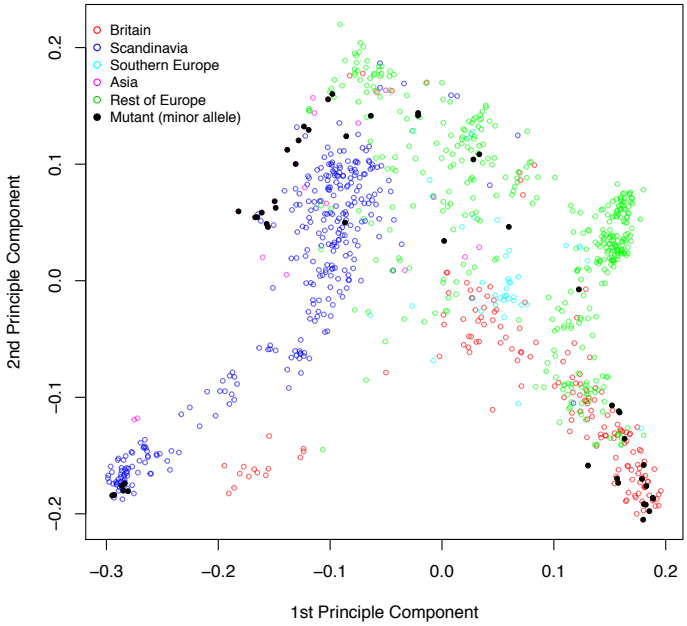
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



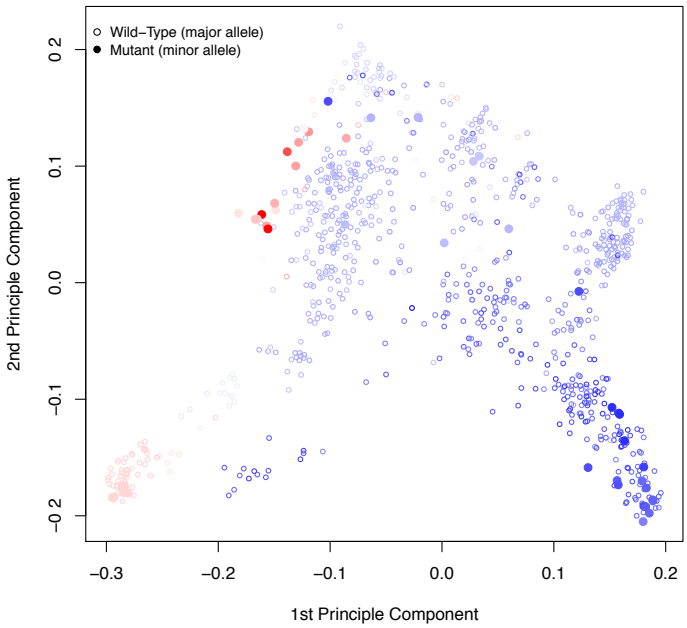
Supplementary Figure 14 - Principle components of the genomic kinship for the two alleles on chromosome 2 at 12169701 bp. Corresponding climate variable: temperature seasonality.

Genomic kinship principle components categorized based on geographical regions.



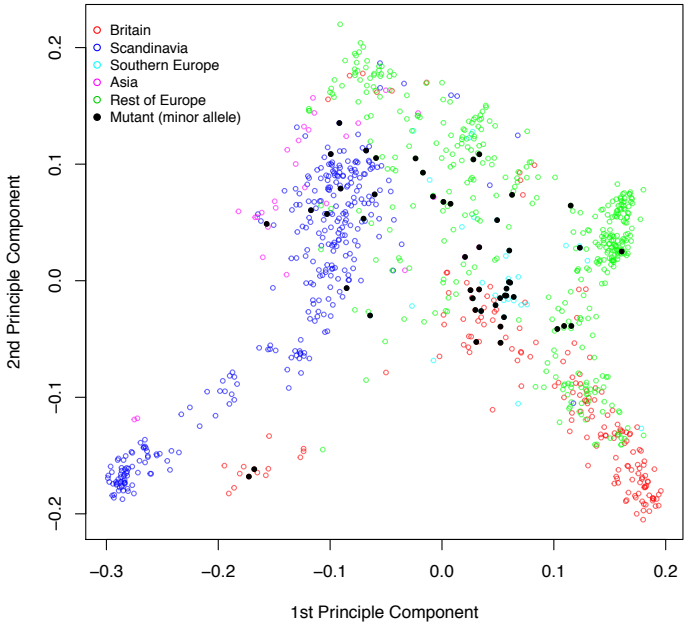
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



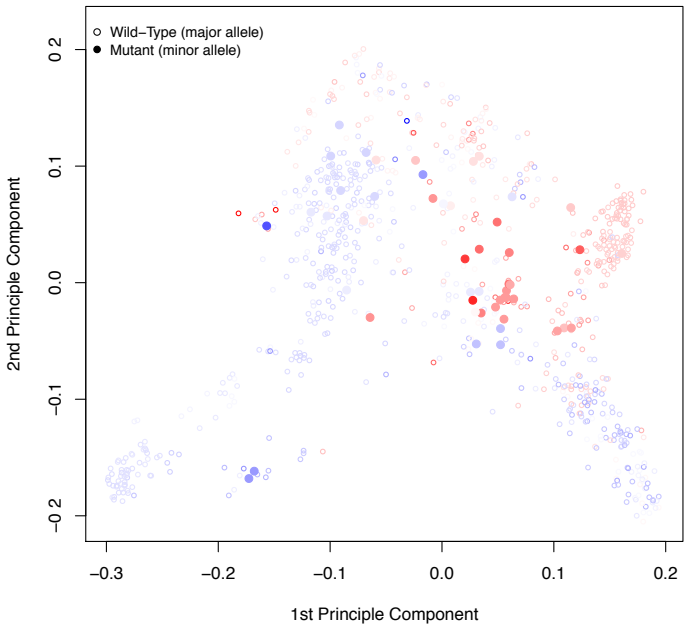
Supplementary Figure 15 - Principle components of the genomic kinship for the two alleles on chromosome 4 at 10406018 bp. Corresponding climate variable: temperature seasonality.

Genomic kinship principle components categorized based on geographical regions.



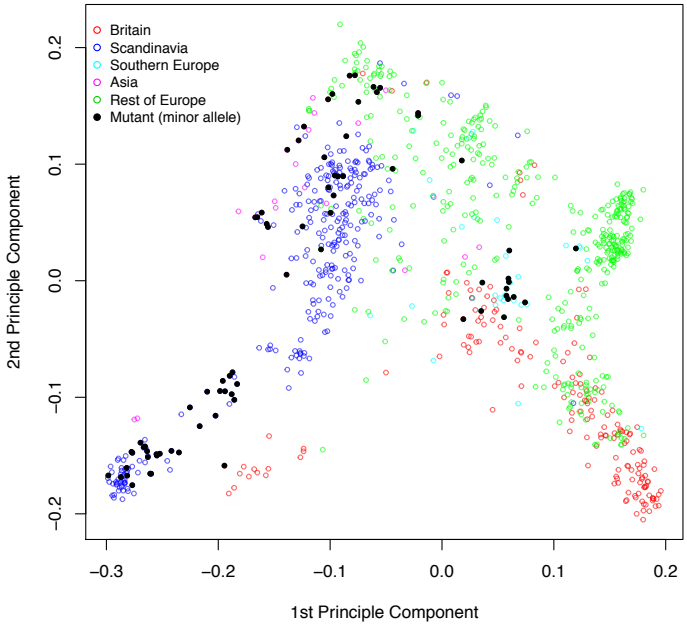
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



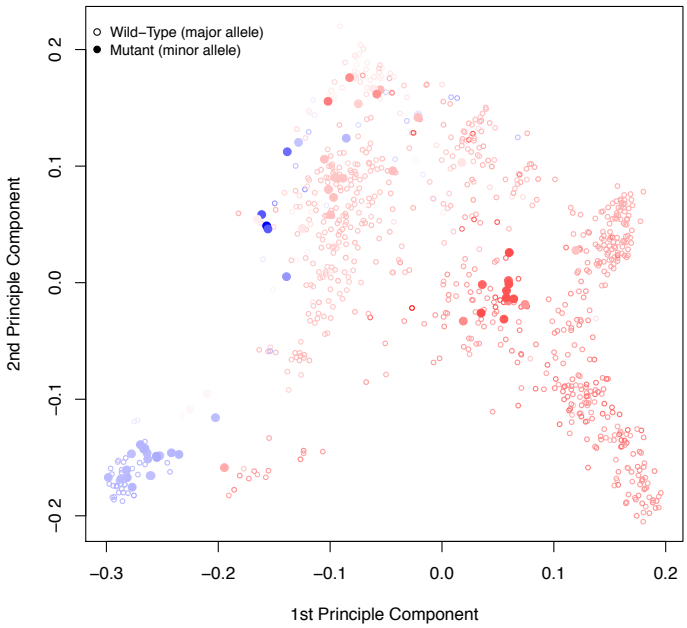
Supplementary Figure 16 - Principle components of the genomic kinship for the two alleles on chromosome 1 at 6936457 bp. Corresponding climate variable: maximum temperature in the warmest month.

Genomic kinship principle components categorized based on geographical regions.



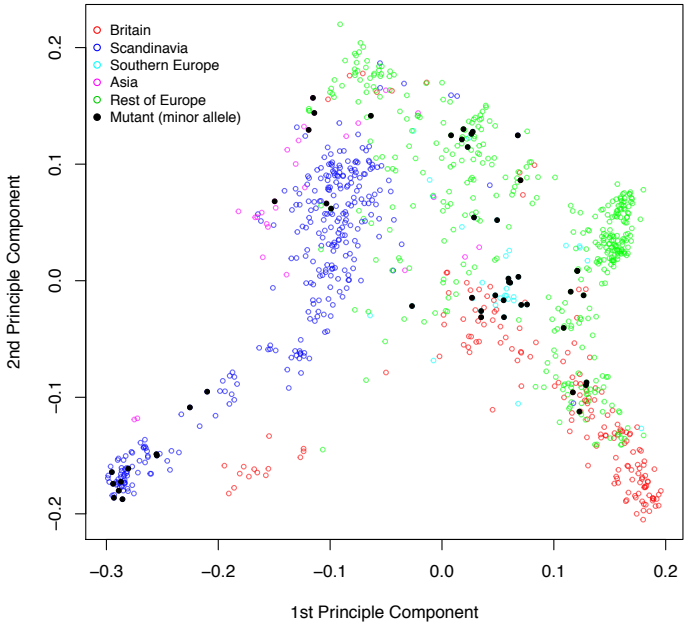
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



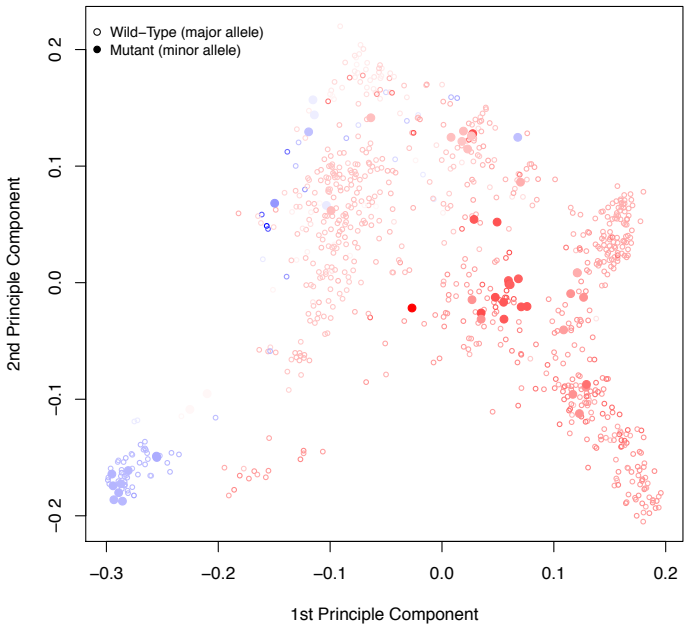
Supplementary Figure 17 - Principle components of the genomic kinship for the two alleles on chromosome 2 at 18620697 bp. Corresponding climate variable: minimum temperature in the coldest month.

Genomic kinship principle components categorized based on geographical regions.



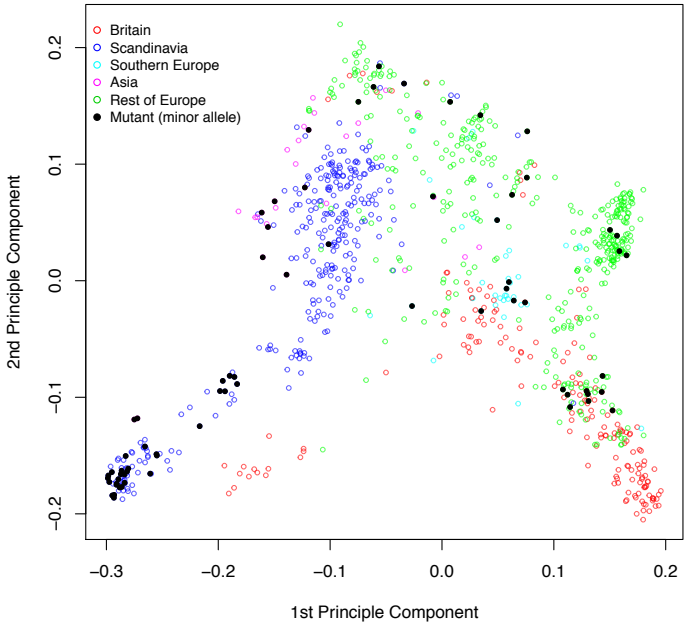
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



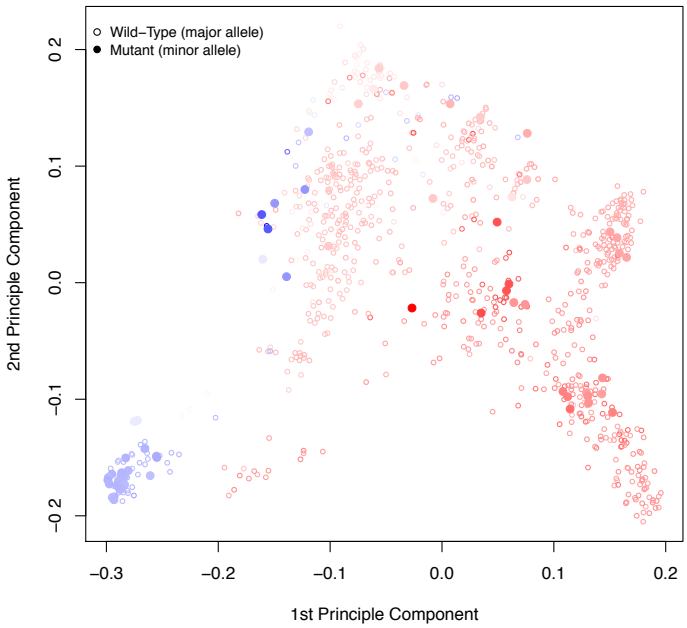
Supplementary Figure 18 - Principle components of the genomic kinship for the two alleles on chromosome 2 at 19397389 bp. Corresponding climate variable: minimum temperature in the coldest month.

Genomic kinship principle components categorized based on geographical regions.



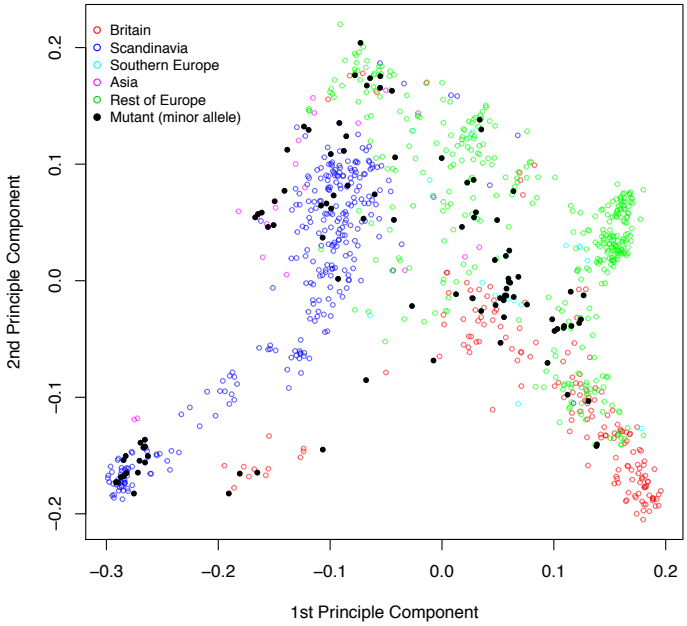
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



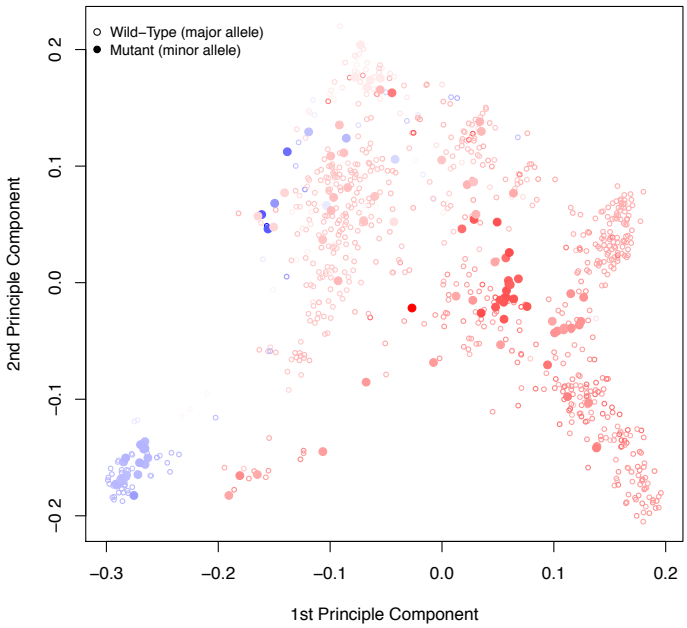
Supplementary Figure 19 - Principle components of the genomic kinship for the two alleles on chromosome 5 at 14067526 bp. Corresponding climate variable: minimum temperature in the coldest month.

Genomic kinship principle components categorized based on geographical regions.



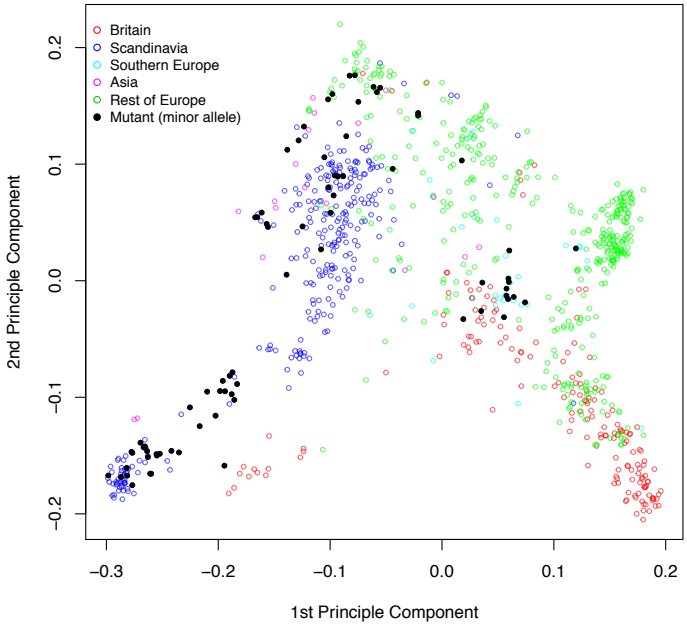
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



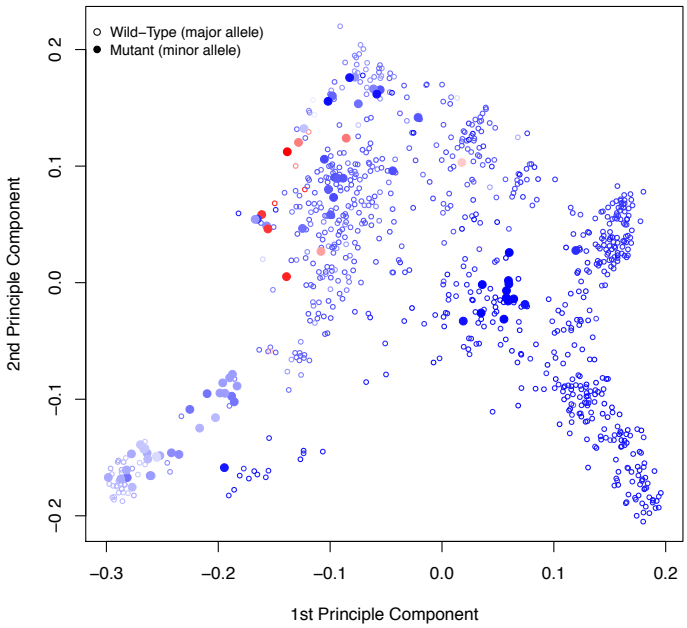
Supplementary Figure 20 - Principle components of the genomic kinship for the two alleles on chromosome 5 at 18397418 bp. Corresponding climate variable: minimum temperature in the coldest month.

Genomic kinship principle components categorized based on geographical regions.



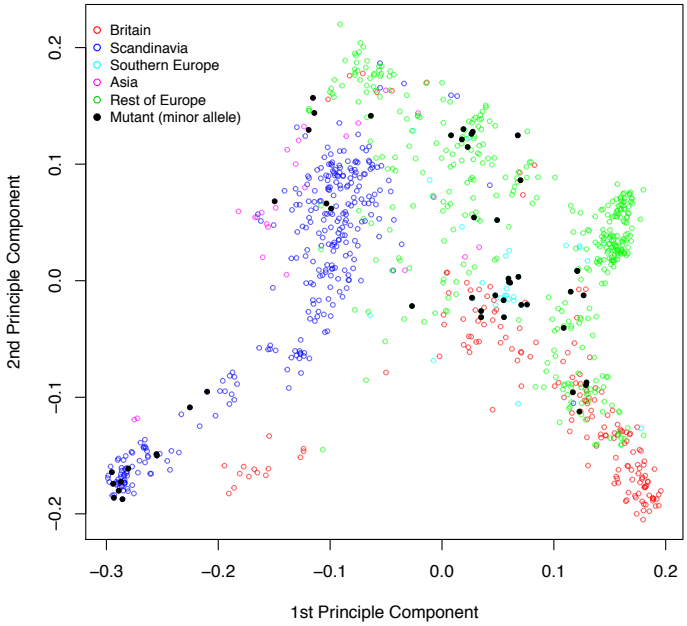
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



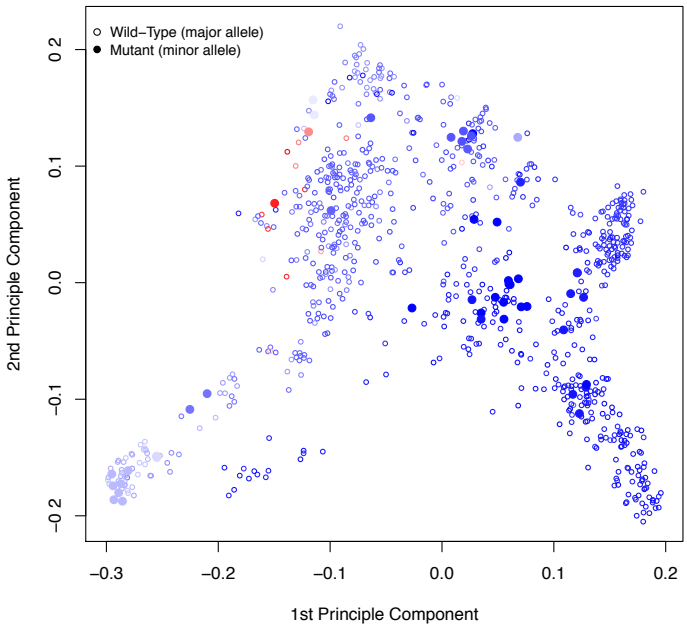
Supplementary Figure 21 - Principle components of the genomic kinship for the two alleles on chromosome 2 at 18620697 bp. Corresponding climate variable: number of consecutive cold days.

Genomic kinship principle components categorized based on geographical regions.



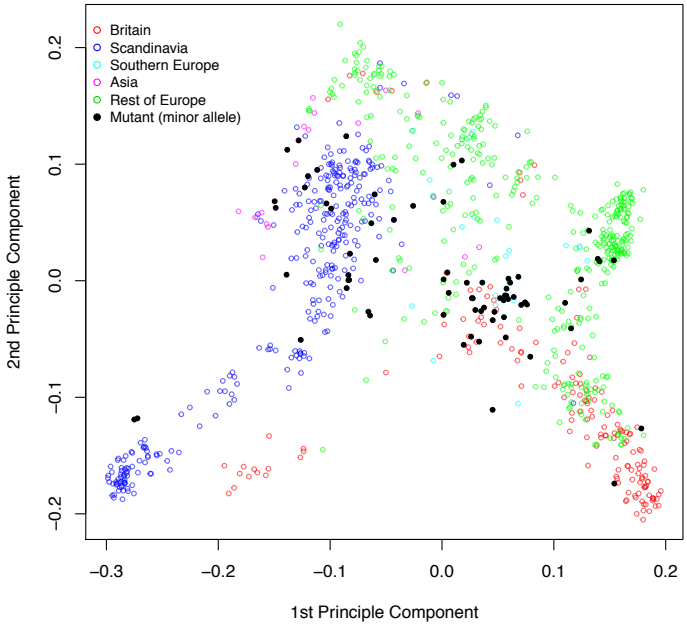
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



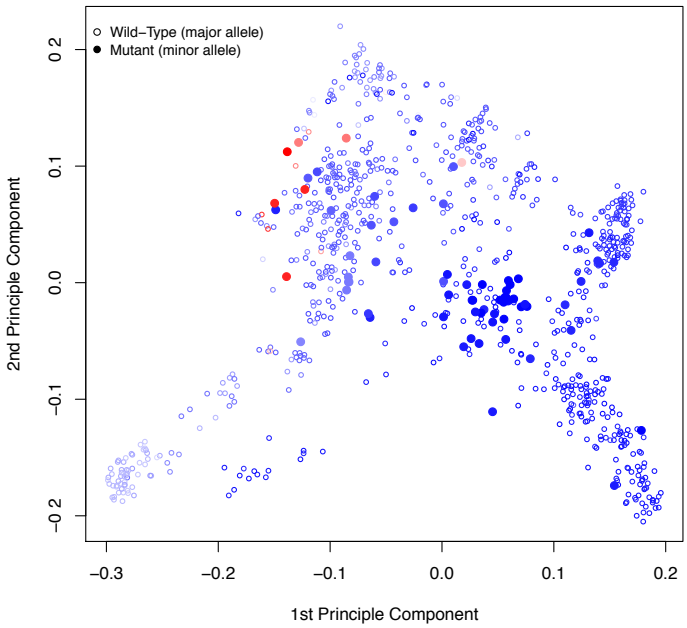
Supplementary Figure 22 - Principle components of the genomic kinship for the two alleles on chromosome 2 at 19397389 bp. Corresponding climate variable: number of consecutive cold days.

Genomic kinship principle components categorized based on geographical regions.



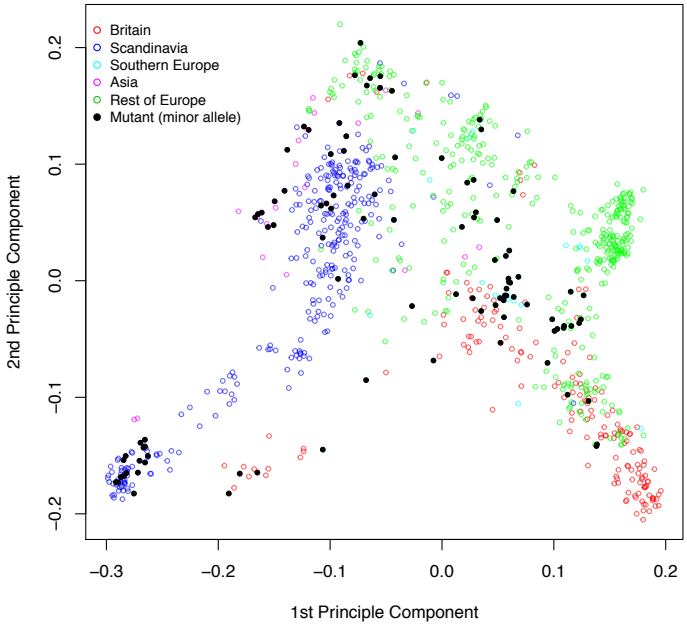
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



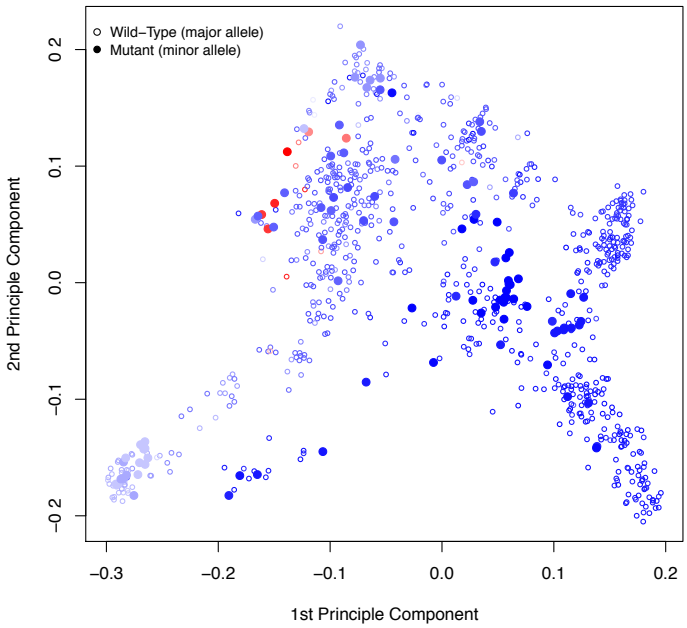
Supplementary Figure 23 - Principle components of the genomic kinship for the two alleles on chromosome 5 at 7492277 bp. Corresponding climate variable: number of consecutive cold days.

Genomic kinship principle components categorized based on geographical regions.



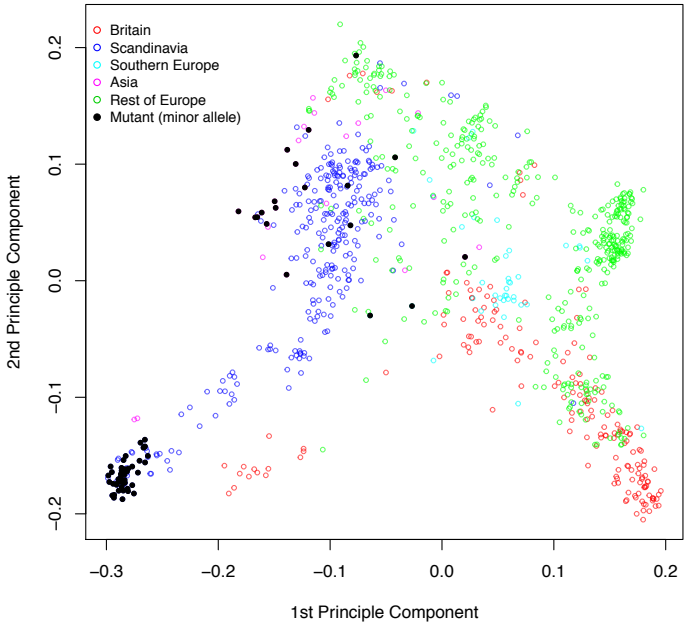
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



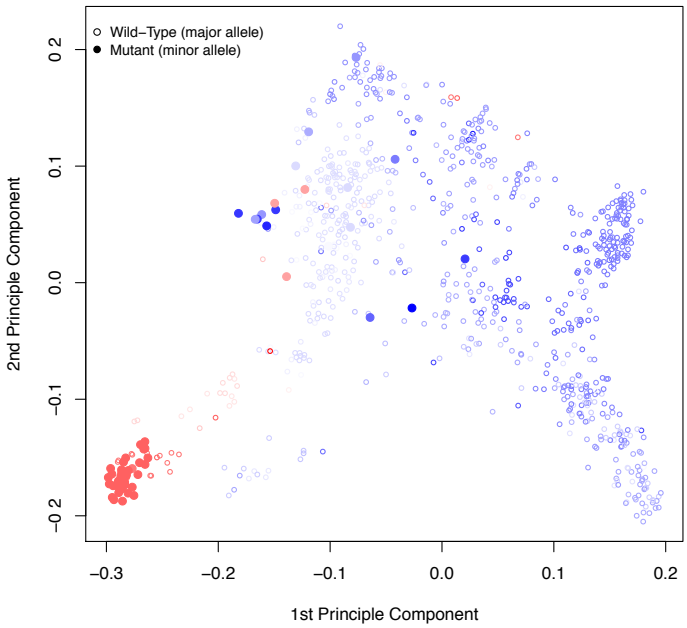
Supplementary Figure 24 - Principle components of the genomic kinship for the two alleles on chromosome 5 at 18397418 bp. Corresponding climate variable: number of consecutive cold days.

Genomic kinship principle components categorized based on geographical regions.



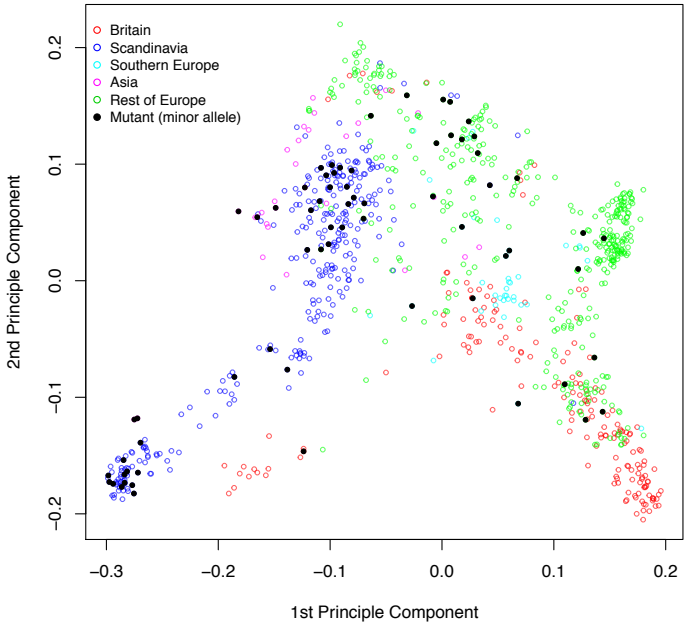
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



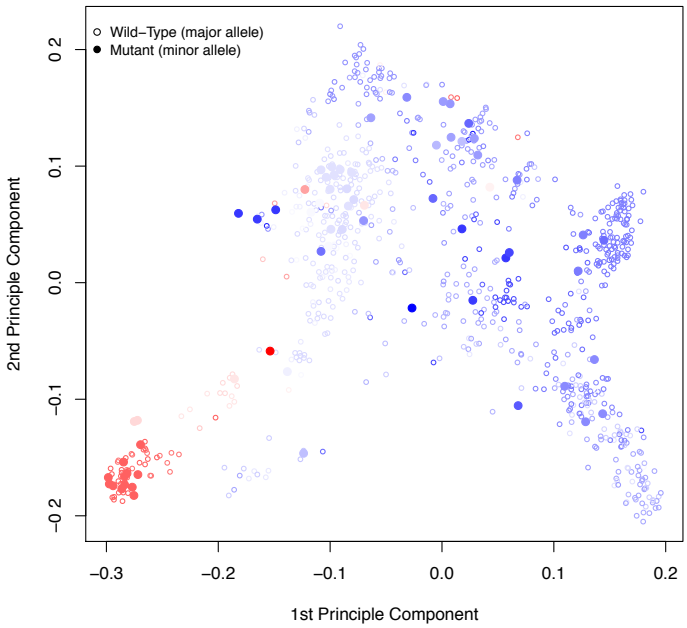
Supplementary Figure 25 - Principle components of the genomic kinship for the two alleles on chromosome 2 at 12169701 bp. Corresponding climate variable: day length in spring.

Genomic kinship principle components categorized based on geographical regions.



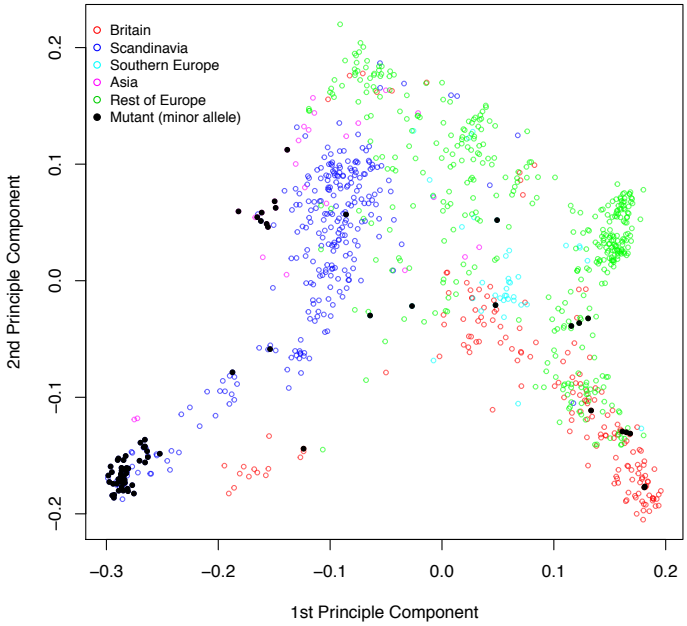
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



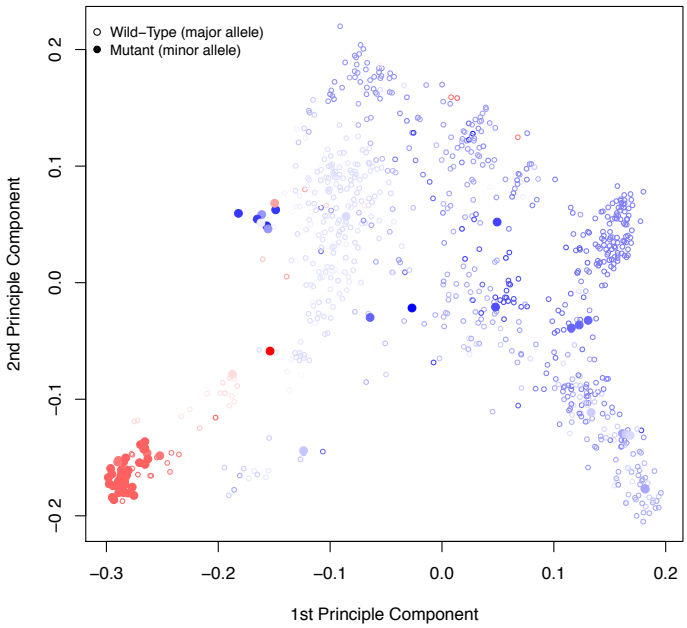
Supplementary Figure 26 - Principle components of the genomic kinship for the two alleles on chromosome 3 at 12642006 bp. Corresponding climate variable: day length in spring.

Genomic kinship principle components categorized based on geographical regions.



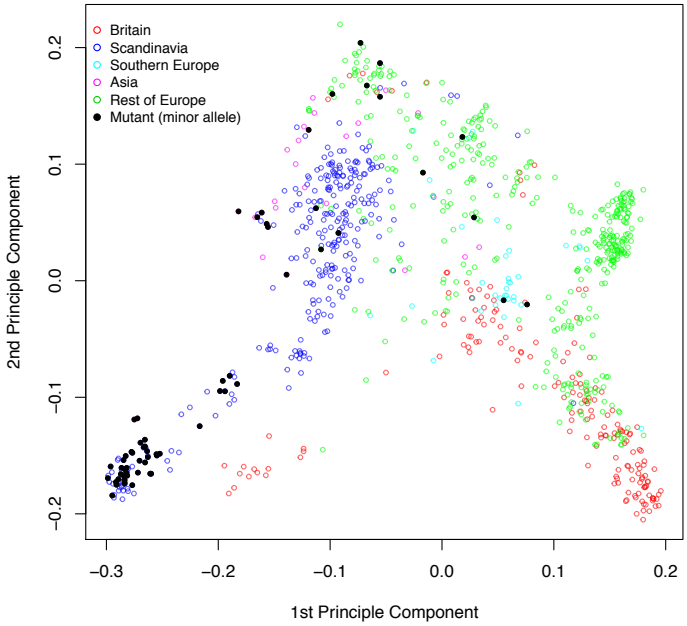
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



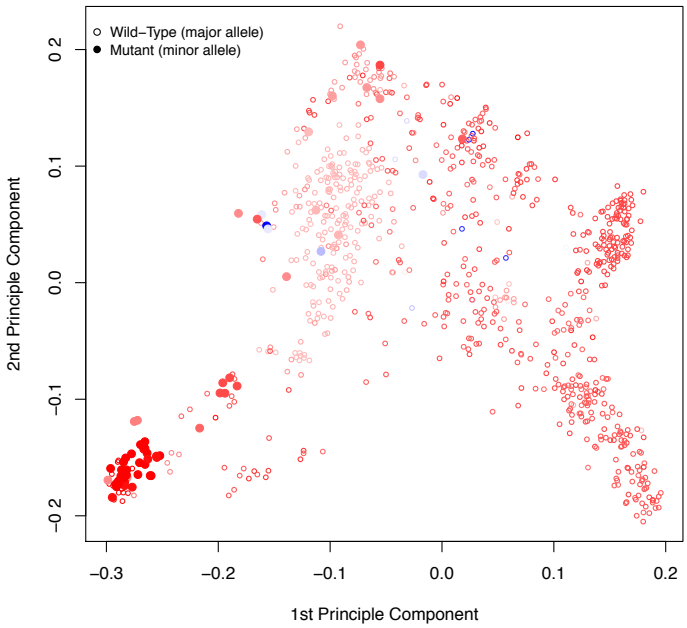
Supplementary Figure 27 - Principle components of the genomic kinship for the two alleles on chromosome 4 at 14788320 bp. Corresponding climate variable: day length in spring.

Genomic kinship principle components categorized based on geographical regions.



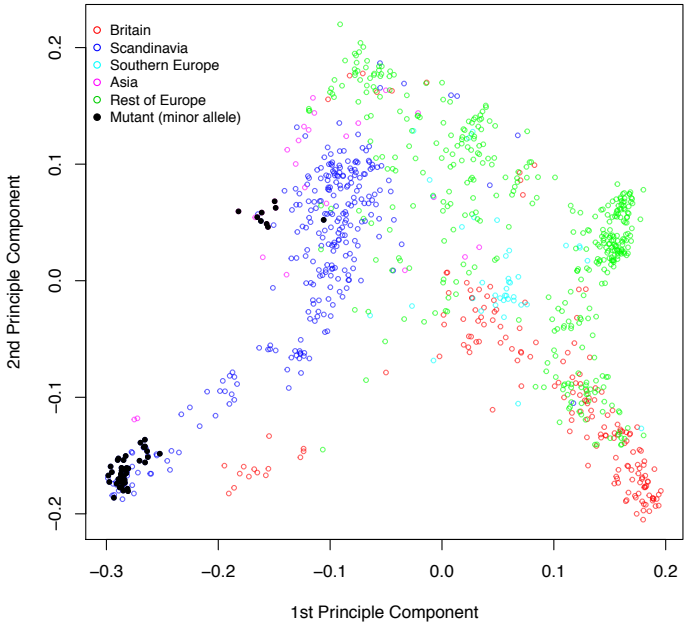
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



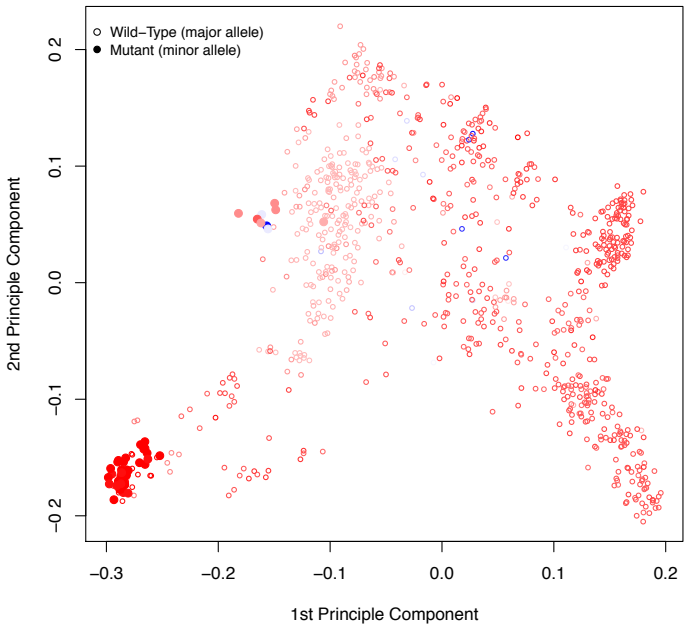
Supplementary Figure 28 - Principle components of the genomic kinship for the two alleles on chromosome 3 at 1816353 bp. Corresponding climate variable: relative humidity in spring.

Genomic kinship principle components categorized based on geographical regions.



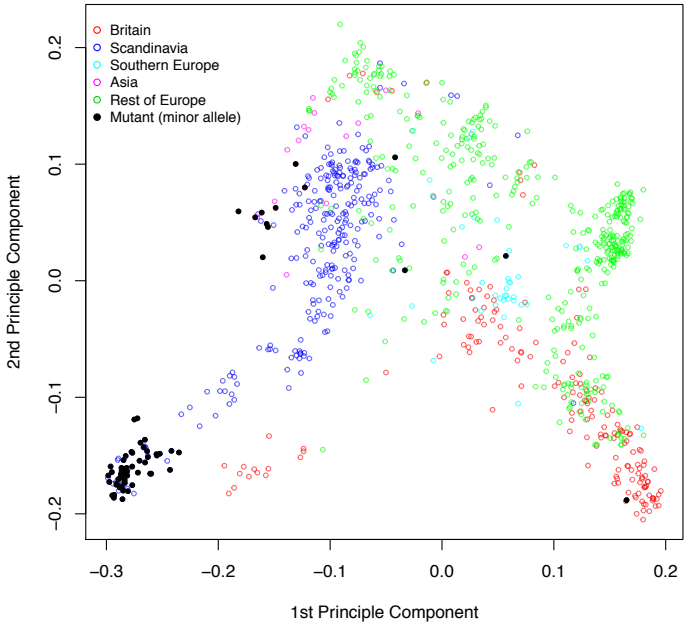
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



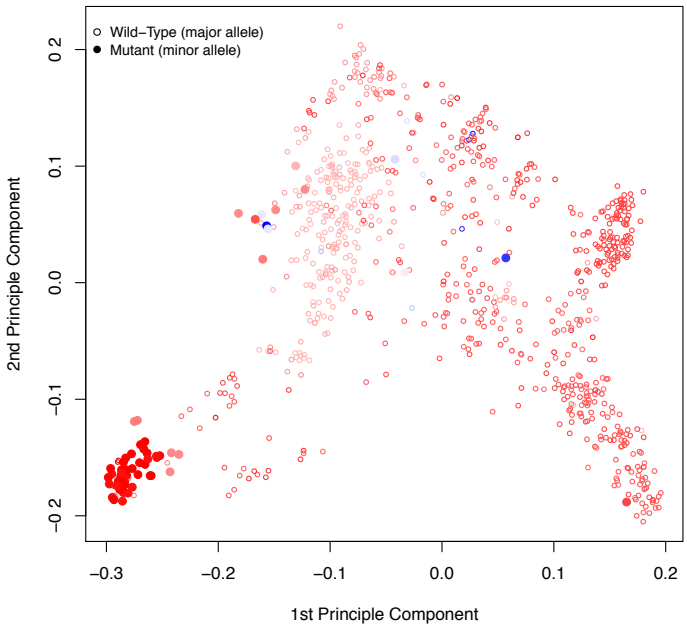
Supplementary Figure 29 - Principle components of the genomic kinship for the two alleles on chromosome 4 at 14834441 bp. Corresponding climate variable: relative humidity in spring.

Genomic kinship principle components categorized based on geographical regions.



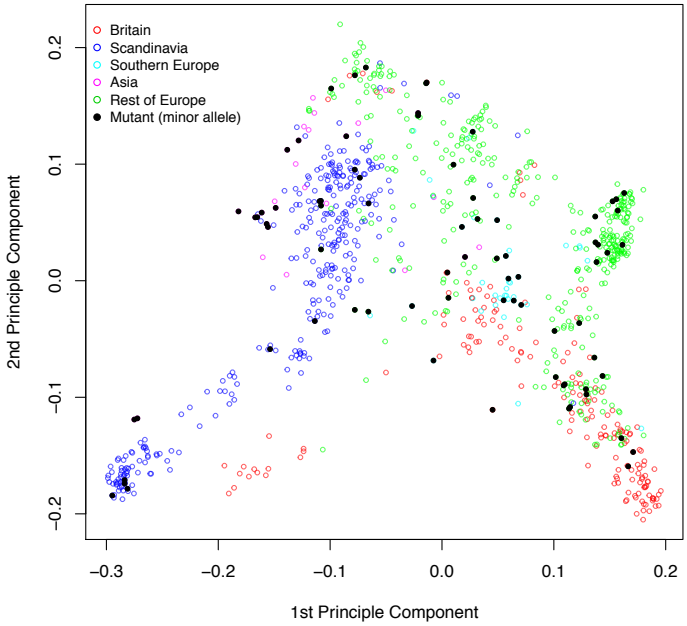
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



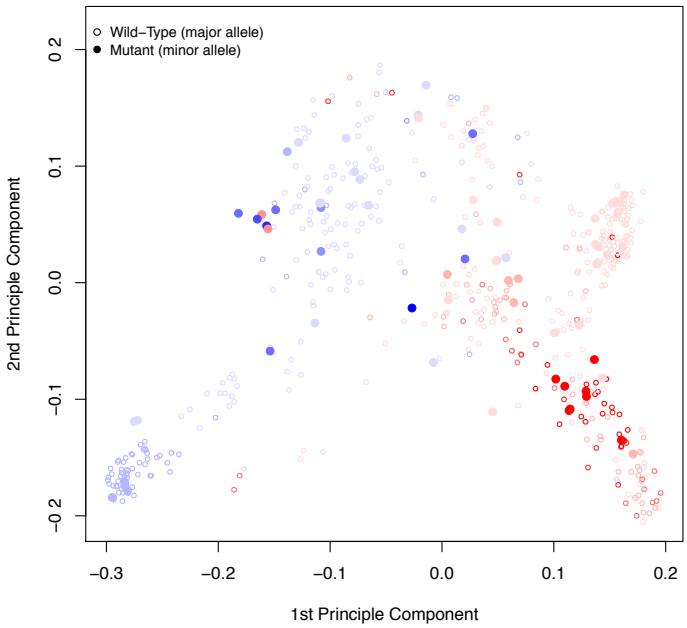
Supplementary Figure 30 - Principle components of the genomic kinship for the two alleles on chromosome 5 at 8380640 bp. Corresponding climate variable: relative humidity in spring.

Genomic kinship principle components categorized based on geographical regions.



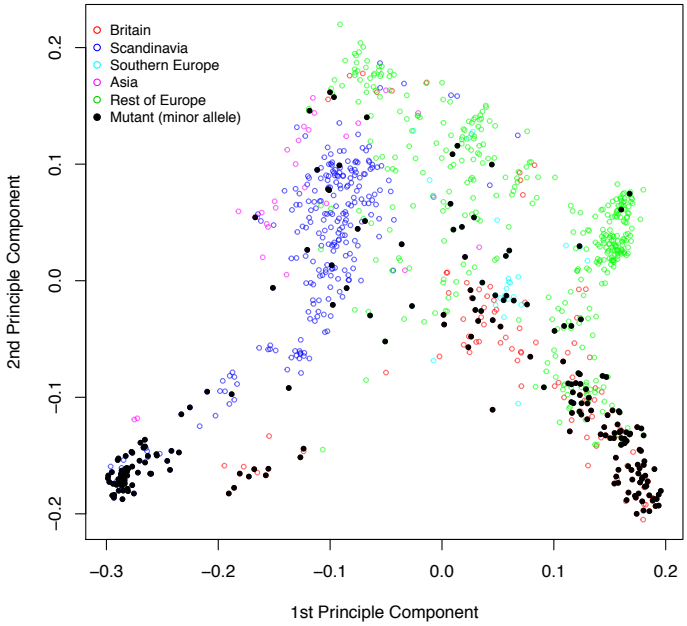
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



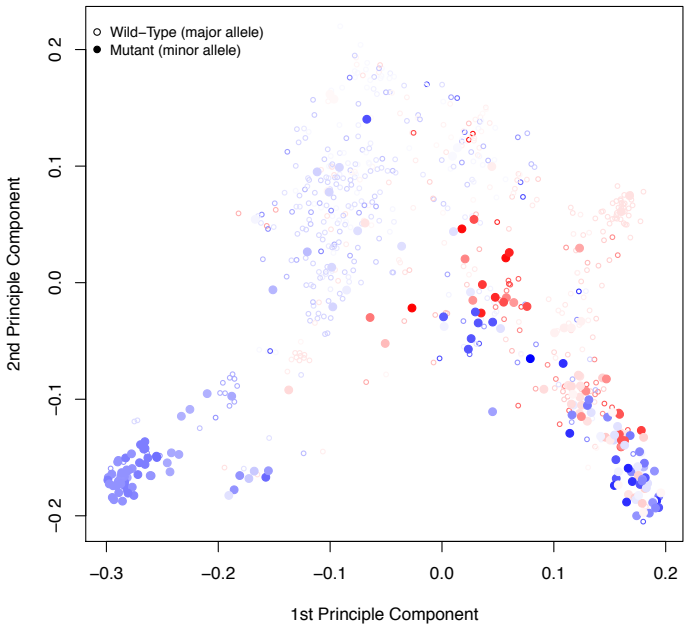
Supplementary Figure 31 - Principle components of the genomic kinship for the two alleles on chromosome 3 at 576148 bp. Corresponding climate variable: length of the growing season.

Genomic kinship principle components categorized based on geographical regions.



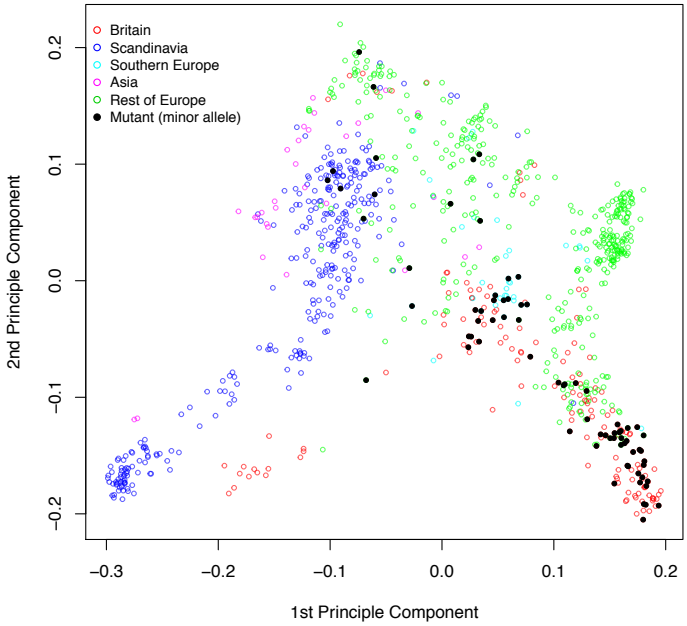
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



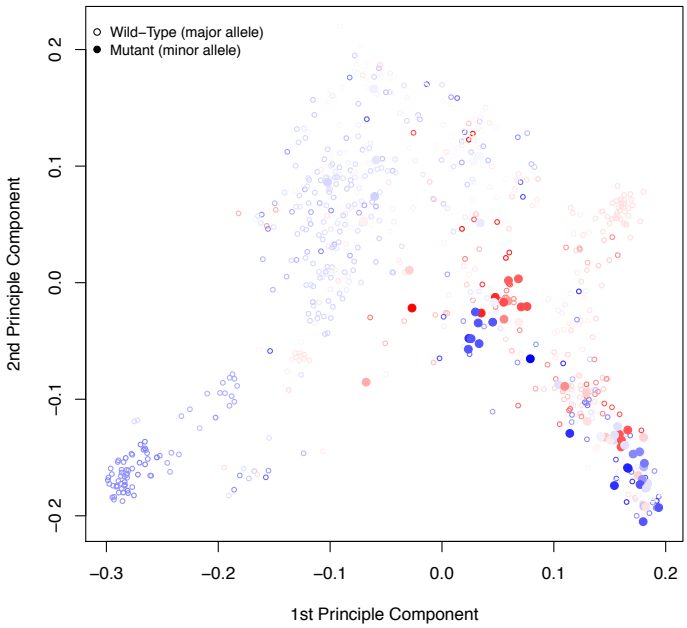
Supplementary Figure 32 - Principle components of the genomic kinship for the two alleles on chromosome 1 at 953031 bp. Corresponding climate variable: number of consecutive frost-free days.

Genomic kinship principle components categorized based on geographical regions.



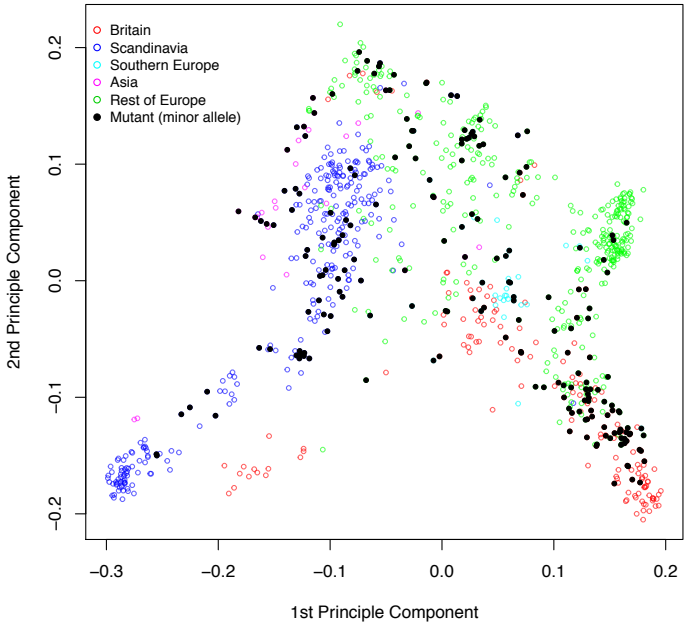
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



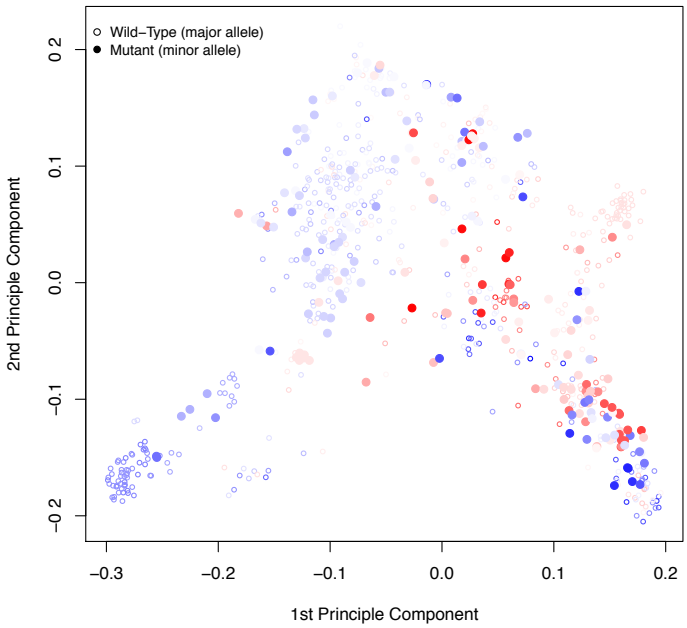
Supplementary Figure 33 - Principle components of the genomic kinship for the two alleles on chromosome 1 at 6463065 bp. Corresponding climate variable: number of consecutive frost-free days.

Genomic kinship principle components categorized based on geographical regions.



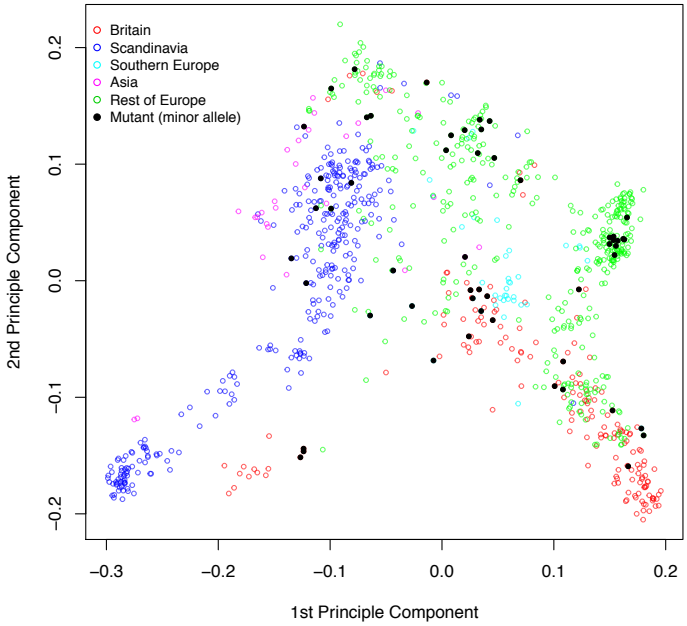
Genomic kinship principle components colored based on the scale of the climate variable.

The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



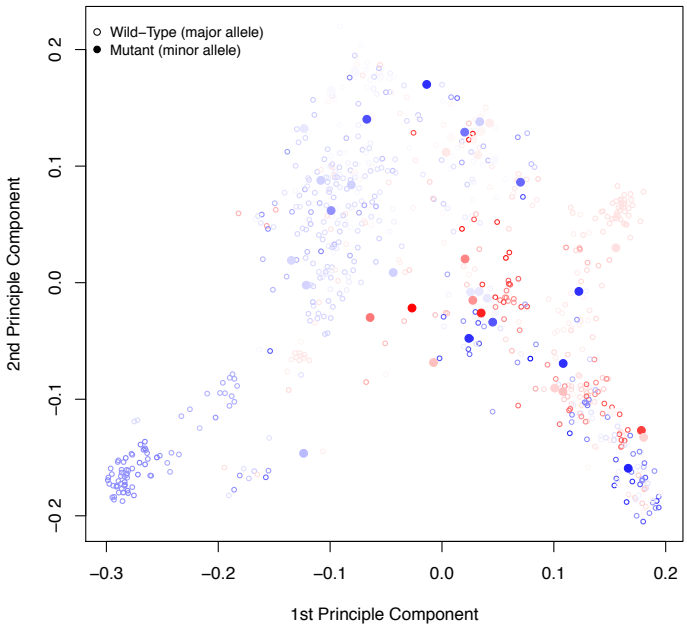
Supplementary Figure 34 - Principle components of the genomic kinship for the two alleles on chromosome 2 at 9904076 bp. Corresponding climate variable: number of consecutive frost-free days.

Genomic kinship principle components categorized based on geographical regions.

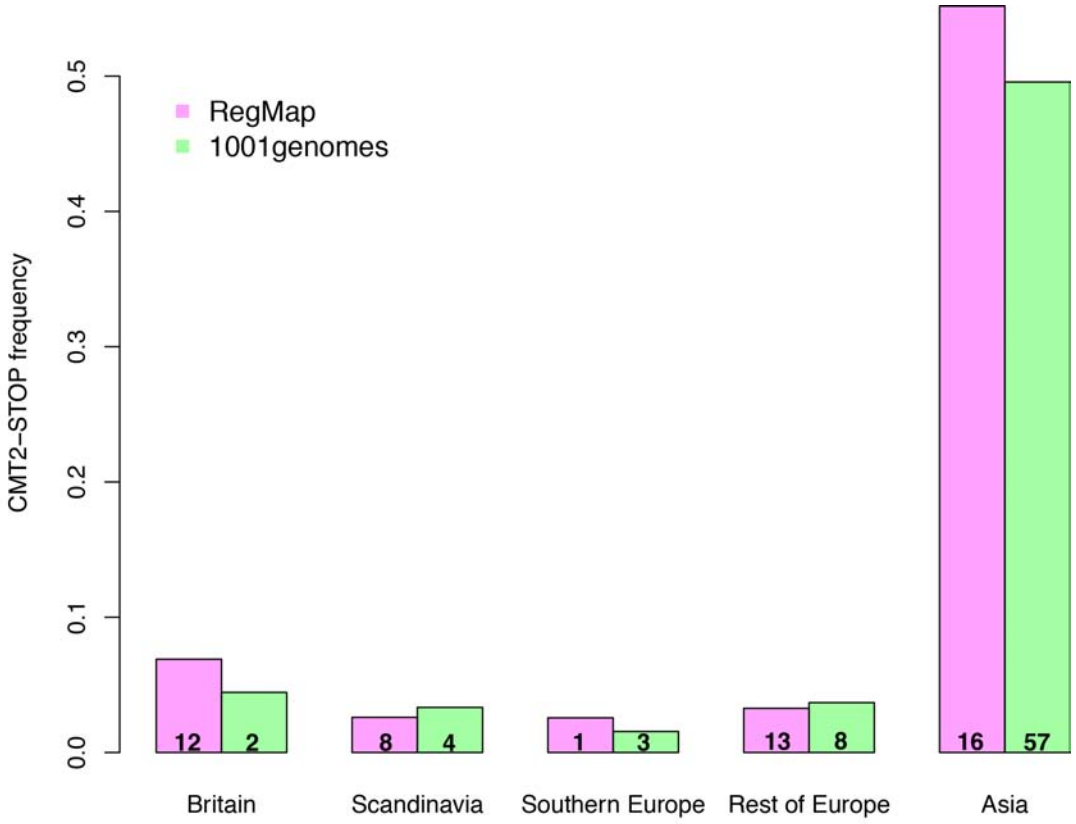


Genomic kinship principle components colored based on the scale of the climate variable.

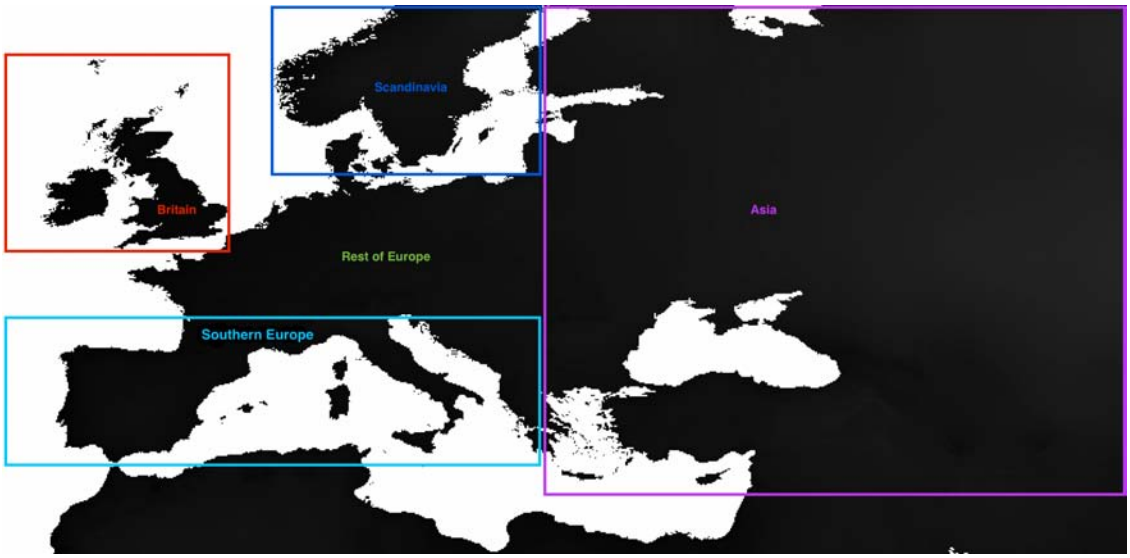
The colors scale from pure blue (the minimum climate variable value) to pure red (the maximum value).



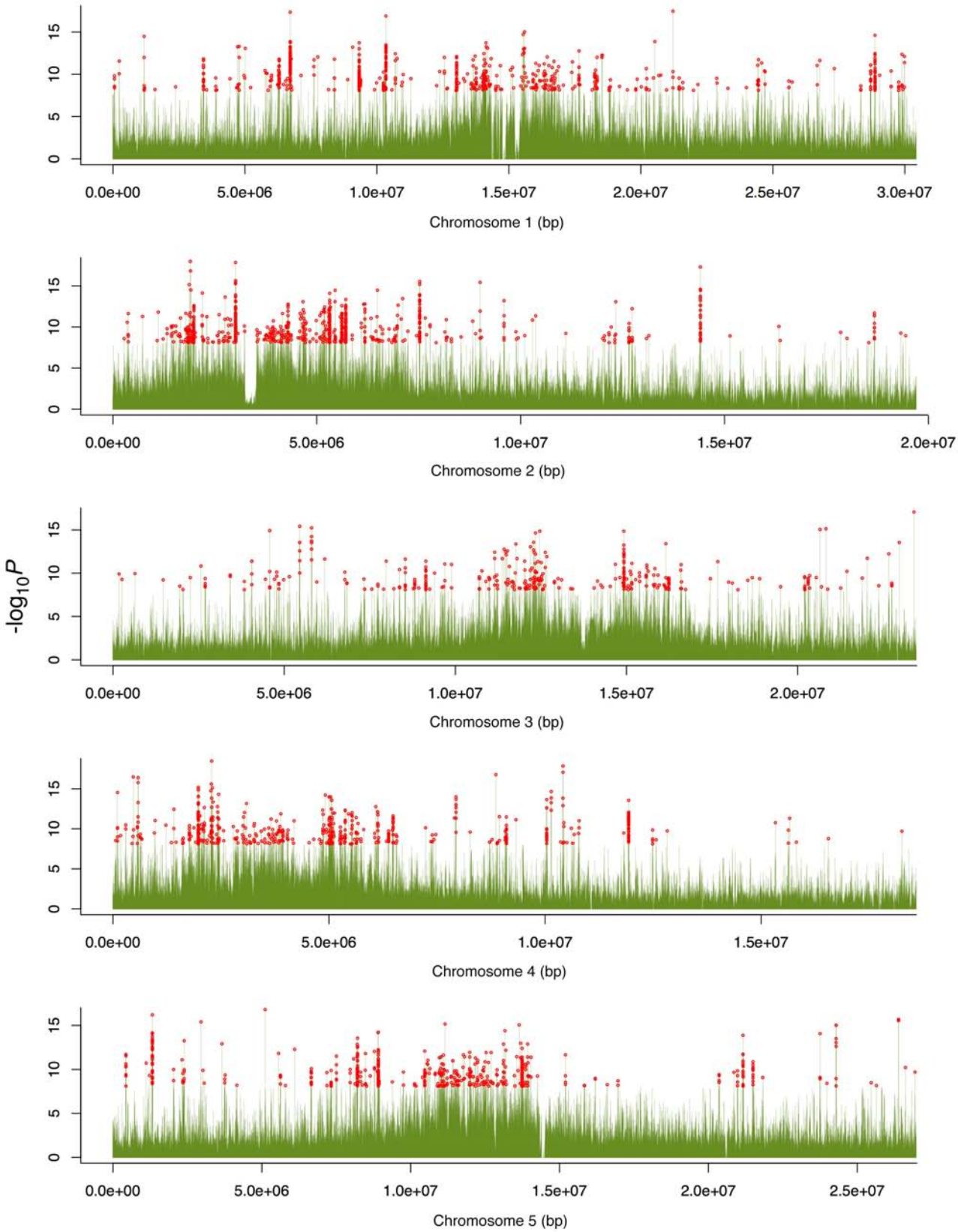
Supplementary Figure 35 - Principle components of the genomic kinship for the two alleles on chromosome 5 at 18061531 bp. Corresponding climate variable: number of consecutive frost-free days.



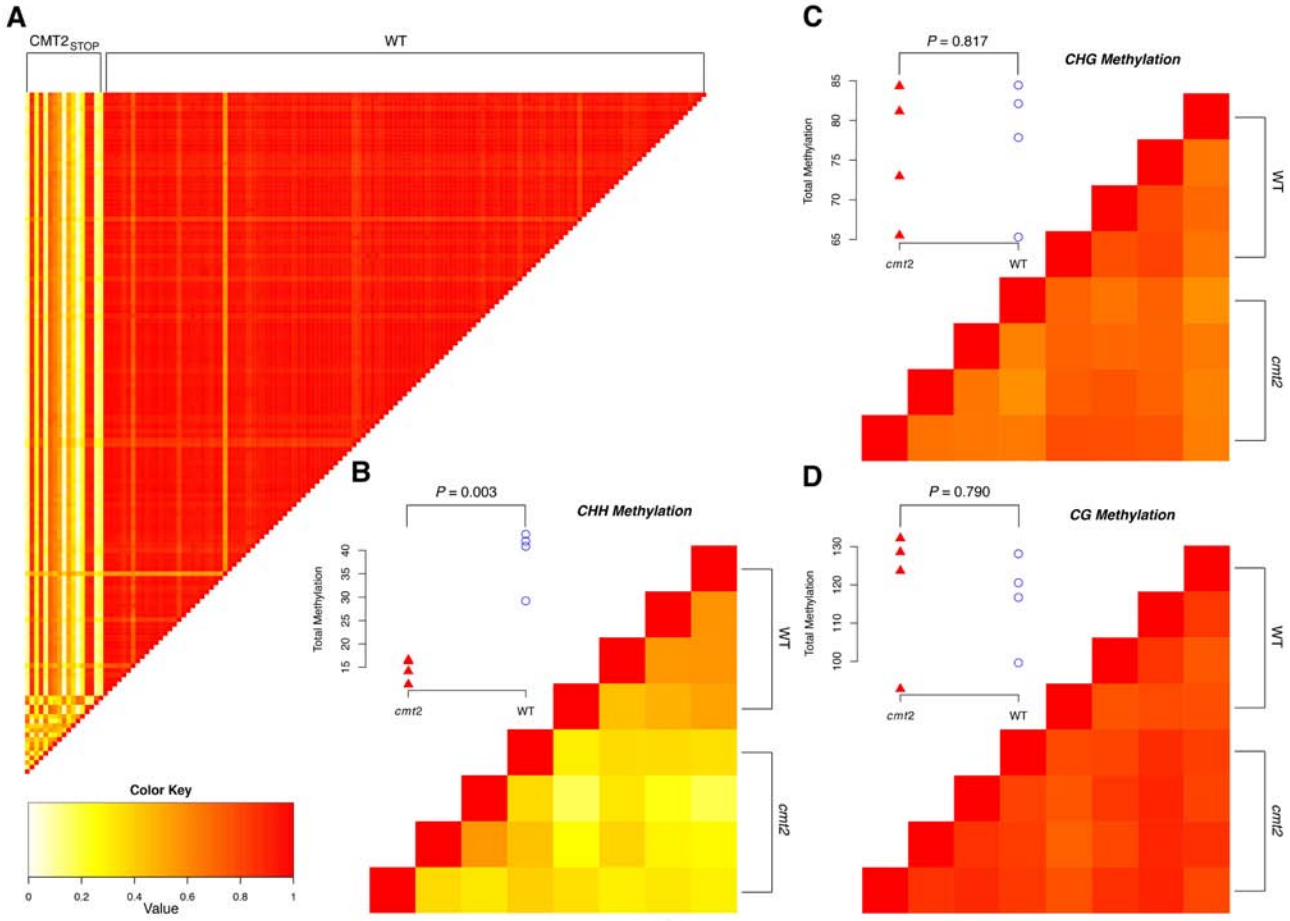
Supplementary Figure 36: Comparison between the RegMap and 1001genomes collections in terms of the allele-frequency of *CMT2$_{STOP}$* across different geographic regions in the Euroasian *A. thaliana* population. The numbers in the bars are the number of *CMT2$_{STOP}$* alleles in this area.



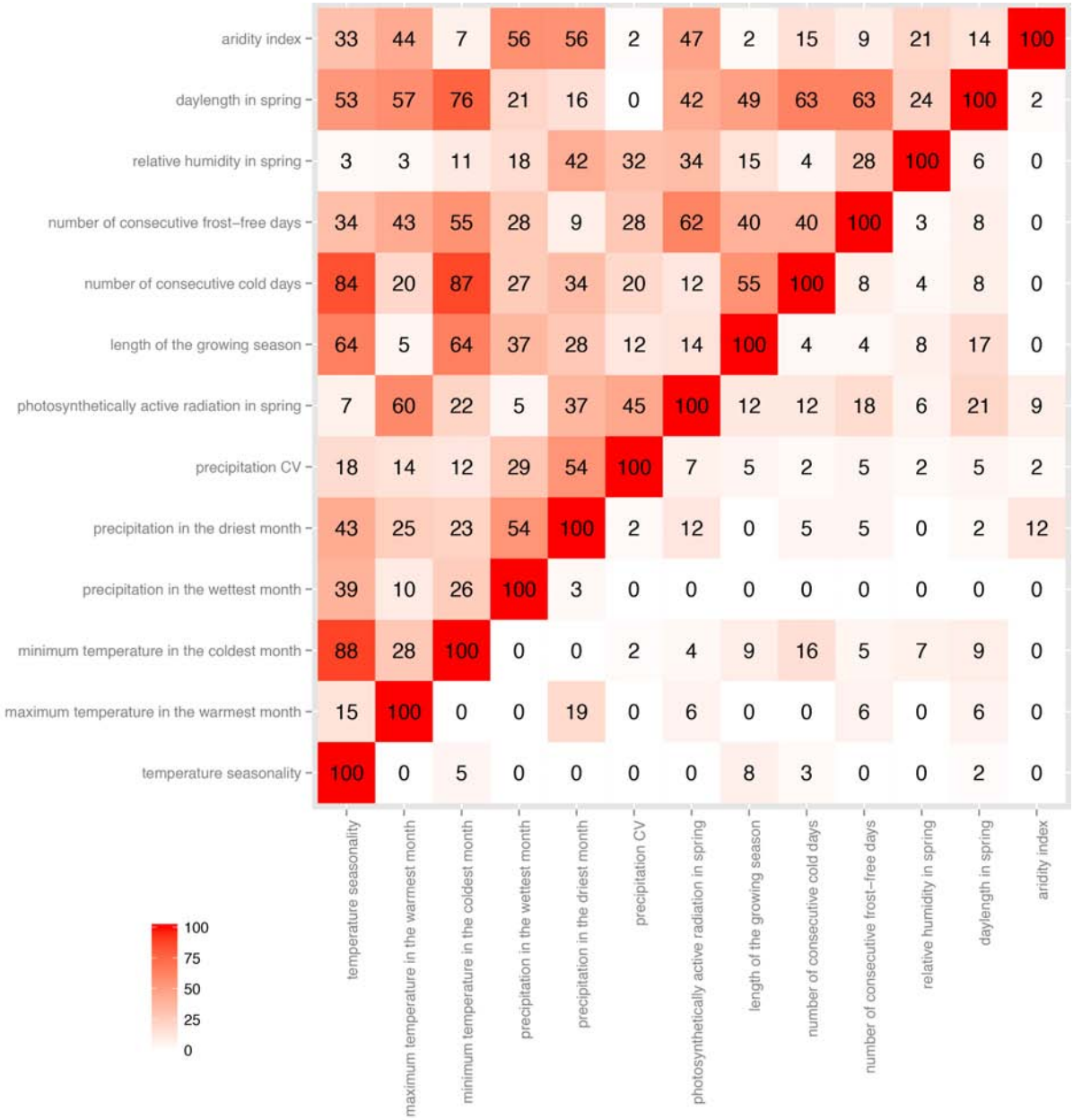
Supplementary Figure 37 - Defined geographical regions across the Euro-Asia sampling area.



Supplementary Figure 38: Methylome-wide association tests for CMT2$_{STOP}$ genotypes. The significant methylation sites passing the Bonferroni-corrected significance threshold are marked in red, which were used in the validation analysis in the CMT2 knockouts.

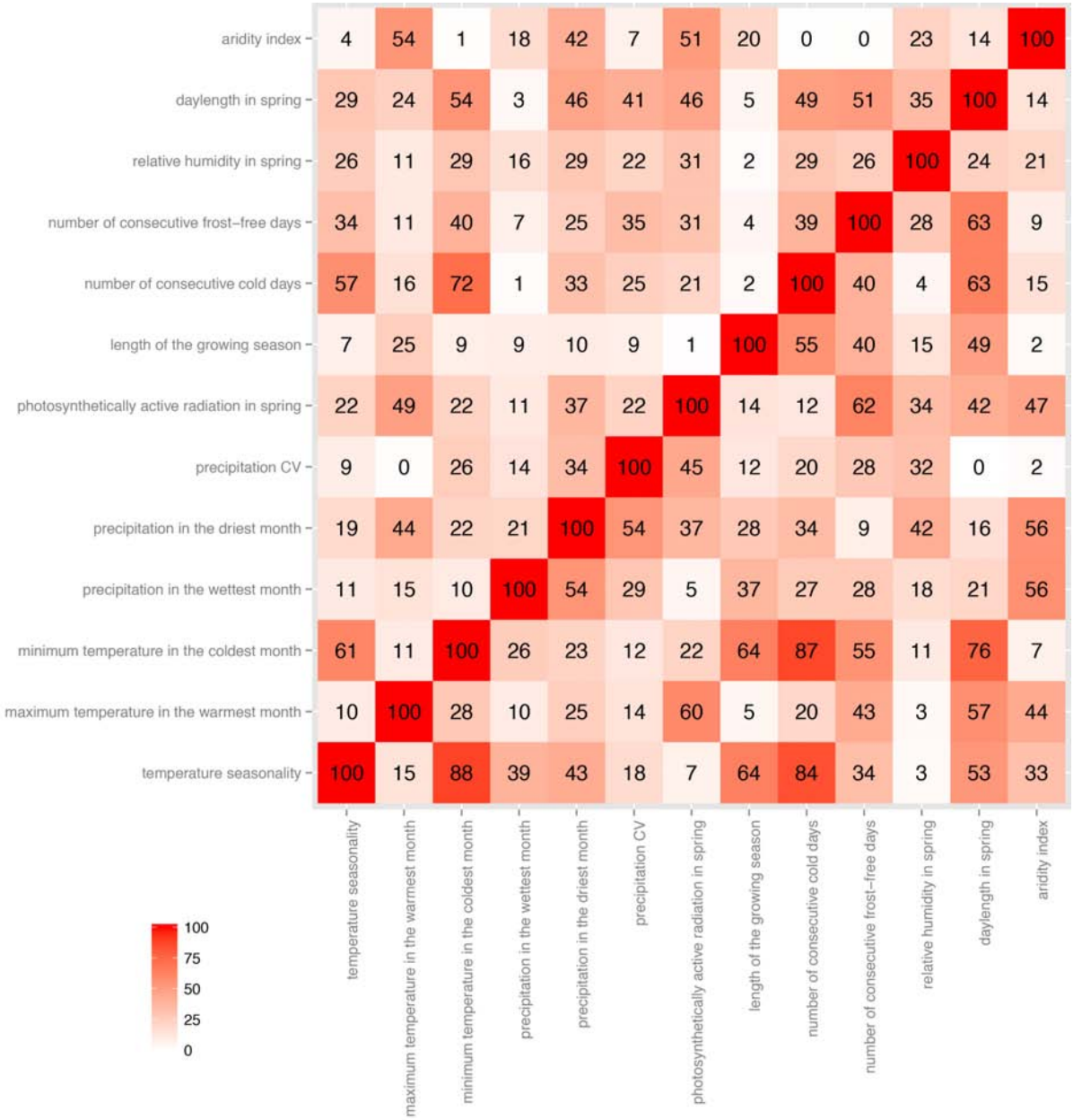


Supplementary Figure 39: Genome-wide methylation patterns for different *CMT2* genotypes. The genome-wide CHH-methylation pattern is similar across the 131 *A. thaliana* CMT2$_{WT}$ **(A)**, but not so among the 17 accessions carrying the loss-of-function (CMT2$_{STOP}$) allele or between the CMT2$_{WT}$ and CMT2$_{STOP}$ accessions. The most divergently CHH methylated sites between natural CMT2$_{WT}$ and *STOP*$_{CMT2}$ accessions are also differentially methylated between CMT2$_{WT}$ and *cmt2* plants, illustrated here by the degree of methylation sharing at these sites amongst four CMT2$_{WT}$ and four *cmt2* mutants (*cmt2-3*, *cmt2-4*, *cmt2-5*, *cmt2-6*) {Zemach: 2013dj}**(B)**. No such differential methylation was found in neither for CHG **(C)** nor CG **(D)** sites. The color key represent identity-by-methylation-state (IBMS) values in (**A**), and correlation coefficients of methylation scores in (**B**, **C**, **D**).



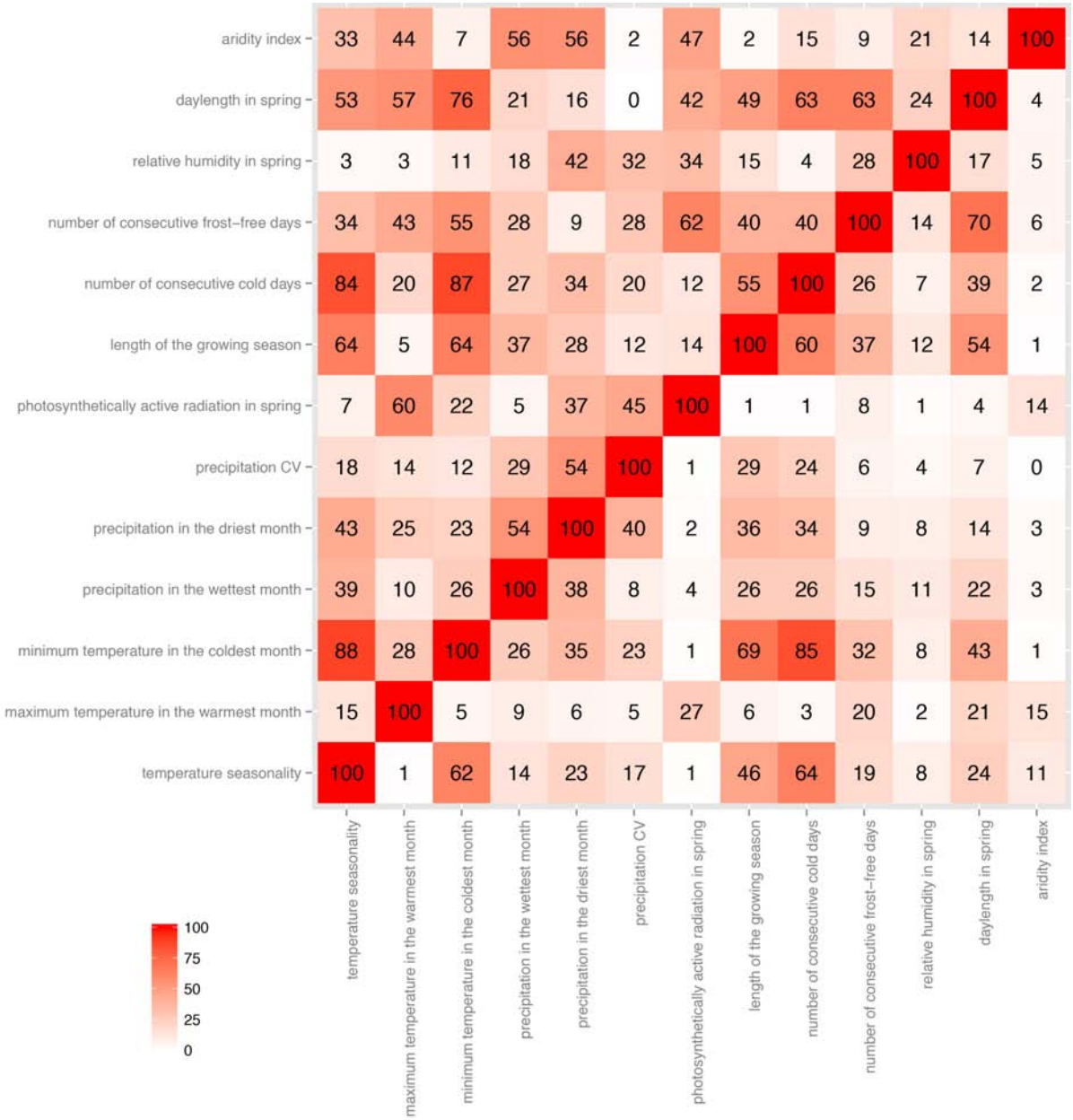
Supplementary Figure 40 - Comparison between the correlations among the climate variables (upper triangle) and the overlap in variance-heterogeneity GWA profiles (lower triangle). Numbers shown in the figure are percentages. Pearson's correlation coefficients were calculated for each pair of the climate variables. Overlaps in GWA profiles were calculated as the proportion of shared SNPs above the threshold of 1.0×10^{-4}.

Supplementary Figure 41 - Comparison between the correlations among the residual climate variables after genomic kinship correction (upper triangle) and the overlap in variance-heterogeneity GWA profiles (lower triangle). Numbers shown in the figure are percentages. Pearson's correlation coefficients were calculated for each pair of the climate variables. Overlaps in GWA profiles were calculated as the proportion of shared SNPs above the threshold of 1.0×10^{-4}.



Supplementary Figure 42 - Comparison between the correlations among the residual climate variables after genomic kinship correction (upper triangle) and the correlations among the original climate variables (lower triangle). Numbers shown in the figure are percentages. Pearson's correlation coefficients were calculated for each pair of the climate variables.

Supplementary Figure 43 - Comparison between the correlations among the climate variables (upper triangle) and the overlap in ordinary GWA profiles (lower triangle). Numbers shown in the figure are percentages. Pearson's correlation coefficients were calculated for each pair of the climate variables. Overlaps in GWA profiles were calculated as the proportion of shared SNPs above the threshold of 1.0×10^{-4}.



Supplementary Figure 44 - Comparison between the correlations among the climate variables (upper triangle) and the overlap in simple GWA profiles without correction for population structure (lower triangle). Numbers shown in the figure are percentages. Pearson's correlation coefficients were calculated for each pair of the climate variables. Overlaps in GWA profiles were calculated as the proportion of shared SNPs above the threshold of 1.0×10^{-4}.

Table S1: Mis- and non-sense mutations in high-LD with genome-wide significant, non-additive associations to climate adaptability.

Trait	Chrom	Pos (bp)	Locus	Gene name	Consequence	Protein position	AA Change	Codon change	MAF	Mutant analysis	
										PASE	MSA
Temperature seasonality											
	4	10 405 599	AT4G19000	IWS2	missense	33	S/T	Tcg/Acg	0.11	0.21	0.35
	4	10 410 039	AT4G19006		missense	240	H/L	cAc/cTc	0.11	0.52	0.41
	4	10 411 801	AT4G19010		missense	539	N/H	Aac/Cac	0.11	0.29	0.08
	4	10 412 877	AT4G19010		missense	373	S/L	tCg/tTg	0.11	0.52	0.14
	4	10 413 290	AT4G19010		missense	311	A/V	gCa/gTa	0.12	0.35	0.68
	4	10 413 374	AT4G19010		missense	283	K/R	aAg/aGg	0.12	0.23	0.2
	4	10 414 556	AT4G19020	CMT2	stop	11	E/*	Gag/Tag	0.11	STOP	STOP
	4	10 414 640	AT4G19020	CMT2	frameshift	39	-	-	0.00	FS	FS
	4	10 414 640	AT4G19020	CMT2	missense	39	E/K	Gaa/Aaa	0.12	0.63	0.17
	4	10 414 747	AT4G19020	CMT2	missense	74	N/K	aaC/aaG	0.11	0.45	0.11
	4	10 415 805	AT4G19020	CMT2	missense	354	G/D	gGc/gAc	0.11	0.58	0.11
	4	10 415 833	AT4G19020	CMT2	missense	363	L/F	ttA/ttT	0.11	0.2	0.14
	4	10 416 047	AT4G19020	CMT2	missense	435	G/S	Ggt/Agt	0.11	0.26	0.14
	4	10 445 425	AT4G19060		missense	270	I/T	aTt/aCt	0.11	0.37	0.44
	4	10 447 160	AT4G19070		missense	140	S/R	Agt/Cgt	0.11	0.73	0.11
	4	10 450 001	AT4G19090		missense	17	R/G	Aga/Gga	0.11	0.93	0.07
	4	10 455 784	AT4G19110		missense	192	L/F	ttG/ttT	0.10	0.2	0.45
	4	10 455 784	AT4G19110		missense	248	L/F	ttG/ttT	0.10	0.2	0.44
	4	10 455 784	AT4G19110		missense	248	L/F	ttG/ttT	0.10	0.2	0.44
Maximum temperature in the warmest month											
	1	6 936 457	AT1G19990		missense	183	S/F	tCt/tTt	0.09	0.64	0.2
Minimum temperature in the coldest month											
	5	14 067 526	AT5G35930		missense	963	E/Q	Gaa/Caa	0.12	0.3	0.05
Number of consecutive cold days											
	5	7 492 033	AT5G22560		missense	355	E/D	gaG/gaC	0.14	0.26	0.22
	5	7 492 259	AT5G22560		missense	280	R/T	aGa/aCa	0.13	0.63	0.11
	5	7 492 277	AT5G22560		missense	274	N/T	aAt/aCt	0.14	0.27	0.11
	5	7 492 277	AT5G22560		missense	274	N/S	aAt/aGt	0.14	0.3	0.11
Relative humidity in spring & Day length in spring											
	4	14 788 320	AT4G30200	VEL1	missense	236	E/D	gaA/gaT	0.11	0.26	0.06
	4	14 788 320	AT4G30200	VEL1	missense	253	E/D	gaA/gaT	0.11	0.26	0.06
	4	14 829 581	AT4G30290	XTH19	missense	100	I/V	Att/Gtt	0.07	0.14	0.49
Minimum temperature in the coldest month & Number of consecutive cold days											
	2	19 397 389	AT2G47240		missense	620	K/R	aAa/aGa	0.06	0.23	0.13
	2	19 397 389	AT2G47240		missense	620	K/R	aAa/aGa	0.06	0.23	0.13
Temperature seasonality & Day length in spring											
	2	12 169 734	AT2G28470	BGAL8	missense	672	E/D	gaA/gaT	0.11	0.26	0.06
	2	12 169 734	AT2G28470	BGAL8	missense	678	E/D	gaA/gaT	0.11	0.26	0.06
	2	12 169 828	AT2G28470	BGAL8	missense	647	F/Y	tTc/tAc	0.12	0.36	0.72
	2	12 169 828	AT2G28470	BGAL8	missense	641	F/Y	tTc/tAc	0.12	0.36	0.72
Number of consecutive frost-free days											
	1	954 782	AT1G03790	SOM	missense	65	N/I	aAt/aTt	0.25	0.6	0.06
	1	955 189	AT1G03790	SOM	missense	201	P/T	Cct/Act	0.25	0.22	0.17
	1	955 268	AT1G03790	SOM	missense	227	S/C	tCt/tGt	0.25	0.55	0.4

Table S2: Genes located less than 100Kb up- or down-stream of the leading SNP in the Genome-Wide Association analysis and that also are in high linkage disequilibrium with the SNP ($r^2 > 0.8$)

Trait	Gene	
Temperature seasonality		
	AT4G18960	AG K-box region and MADS-box transcription factor family protein
	AT4G18970	GDSL-like Lipase/Acylhydrolase superfamily protein
	AT4G18975	Pentatricopeptide repeat (PPR) superfamily protein
	AT4G18980	AtS40-3
	AT4G18990	XTH29 xyloglucan endotransglucosylase/hydrolase 29
	AT4G19003	VPS25 E2F/DP family winged-helix DNA-binding domain
	AT4G19030	NLM1 NOD26-like major intrinsic protein 1
	AT4G19035	LCR7 low-molecular-weight cysteine-rich 7
	AT4G19038	LCR15 low-molecular-weight cysteine-rich 15
	AT4G19040	EDR2 ENHANCED DISEASE RESISTANCE 2
	AT4G19045	Mob1/phocein family protein
	AT4G19050	NB-ARC domain-containing disease resistance protein
	AT4G19080	unknown protein
	AT4G19095	unknown protein
	AT4G19100	unknown protein
	AT4G19112	CPuORF25 conserved peptide upstream open reading frame 25
	AT4G19120	ERD3 S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
Maximum temperature in the warmest month		
	AT1G19970	ER lumen protein retaining receptor family protein
	AT1G19980	cytomatrix protein-related
	AT1G20000	TAF11b TBP-associated factor 11B
	AT1G20010	TUB5 tubulin beta-5 chain
	AT1G20015	snoRNA
Minimum temperature in the coldest month		
	AT5G35926	Protein with RNI-like/FBD-like domains
Number of consecutive cold days		
	AT5G22555	unknown protein
	AT5G22570	WRKY38 WRKY DNA-binding protein 38
Day length in spring		
	AT3G30859	transposable element gene
	AT3G30867	pseudogene, putative SNF8 protein homolog
Relative humidity in spring & Day length in spring		
	AT4G30240	Syntaxin/t-SNARE family protein
	AT4G30250	P-loop containing nucleoside triphosphate hydrolases superfamily protein
	AT4G30260	Integral membrane Yip1 family protein
	AT4G30270	MERI5B xyloglucan endotransglucosylase/hydrolase 24
	AT4G30280	ATXTH18 xyloglucan endotransglucosylase/hydrolase 18
	AT4G30300	ATNAP15 non-intrinsic ABC protein 15
	AT4G30320	CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) superfamily protein
	AT4G30330	Small nuclear ribonucleoprotein family protein
	AT4G30340	ATDGK7 diacylglycerol kinase 7
Minimum temperature in the coldest month & Number of consecutive cold days		
	AT2G47250	RNA helicase family protein
	AT2G45150	CDS4 cytidinediphosphate diacylglycerol synthase 4
	AT2G45160	HAM1 GRAS family transcription factor
	AT2G45161	unknown protein

	AT2G45170	ATATG8E AUTOPHAGY 8E
	AT5G45380	ATDUR3 solute:sodium symporters;urea transmembrane transporters
	AT5G45390	CLPP4 CLP protease P4
	AT5G45400	RPA70C Replication factor-A protein 1-related
	AT5G45410	unknown protein
Temperature seasonality & Day length in spring		
	AT2G28410	unknown protein
	AT2G28420	Lactoylglutathione lyase / glyoxalase I family protein
	AT2G28426	unknown protein
	AT2G28430	unknown protein
	AT2G28440	proline-rich family protein
	AT2G28450	zinc finger (CCCH-type) family protein
	AT2G28460	Cysteine/Histidine-rich C1 domain family protein
Relative humidity in spring		
	AT3G06019	unknown protein
	AT3G06020	unknown protein
	AT3G06030	ANP3 NPK1-related protein kinase 3
	AT5G24530	DMR6 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
	AT5G24540	BGLU31 beta glucosidase 31
Length of the growing season		
	AT3G02660	Tyrosyl-tRNA synthetase, class Ib, bacterial/mitochondrial
	AT3G02670	Glycine-rich protein family
	AT3G02680	NBS1 nijmegen breakage syndrome 1
	AT3G02690	nodulin MtN21 /EamA-like transporter family protein
Number of consecutive frost-free days		
	AT1G03780	TPX2 targeting protein for XKLP2
	AT1G03800	ERF10 ERF domain protein 10
	AT1G03810	Nucleic acid-binding, OB-fold-like protein
	AT1G03820	unknown protein
	AT1G03830	guanylate-binding family protein
	AT1G18720	unknown protein
	AT1G18730	NDF6 NDH dependent flow 6
	AT1G18735	other RNA
	AT1G18740	unknown protein
	AT1G18745	NcRNA
	AT1G18750	AGL65 AGAMOUS-like 65
	AT2G23250	UGT84B2 UDP-glucosyl transferase 84B2
	AT2G23260	UGT84B1 UDP-glucosyl transferase 84B1
	AT2G23270	unknown protein
	AT2G23290	AtMYB70 myb domain protein 70
	AT5G44740	POLH
	AT5G44750	REV1
	AT5G44760	C2 domain-containing protein
	AT5G44770	Cysteine/Histidine-rich C1 domain family protein