Genome-wide association study in European beech (*Fagus sylvatica* L.) for drought stress traits

Mila Tost[1,2,3] (0009-0001-3460-5994)**, Ourania Grigoriadou-Zormpa**[2,3] (0009-0000-2994-0643)**, Selina Wilhelmi**[2,3] (0000-0001-7640-6523)**, Timothy Beissinger**[1,3,6] (0000-0002-2882-4074)**, Markus Müller**[2,3] (0000-0001-9990-0719)**, Henning Wildhagen**[4] (0000-0003-4970-8539)**, Alexandru Lucian Curtu**[5] **(0000-0001-8509-279X)**, **Oliver Gailing**[2,3] (0000-0002-4572-2408)

[1)] Division of Plant Breeding Methodology, University of Göttingen, Carl-Sprengel-Weg 1, 37075 Göttingen (Germany).
[2)] Forest Genetics and Forest Tree Breeding, University of Göttingen, Büsgenweg 2, 37077 Göttingen (Germany).
[3)] Center for Integrated Breeding Research (CiBreed), University of Göttingen,
Von-Siebold-Str. 4, 37075 Göttingen (Germany).
[4)] Faculty of Resource Management, HAWK, Büsgenweg 1a, 37077 Göttingen (Germany).
[5)] Department of Silviculture, Transilvania University of Brasov, Sirul Beethoven -1, Bra□ov 500123 (Romania).
[6)] Heritable Agriculture, Mountain View, California, USA, 94040 (USA).

[*]**Corresponding author:** Oliver Gailing (ogailin@gwdg.de)

Running title: GWAS European beech

Key words: GWAS, *Fagus sylvatica*, natural population, population genetics, quantitative genetics, polygenic selection

Abstract

Forest tree breeding is an extremely long and tedious process. To study the genetic architecture of polygenic traits in long-lived species such as forest trees, costly field experiments are implemented. Phenotypic data on traits, that are measured at maturity, are only available after a long time and juvenile-mature correlations are often unknown. Genome-wide association studies (GWAS) aim to identify loci associated with drought stress, tree physiology, growth or wood quality traits, which could be prioritized in breeding programs. Genotypic and phenotypic data were collected from approximately 100 adult beech trees per stand in five locations in the South-Eastern Romanian Carpathians along an altitudinal gradient associated with precipitation and temperature. We performed GWAS using PLINK to identify SNP markers associated with traits related to drought stress. A total of 121 markers on eight chromosomes were identified as being associated with stomatal density. Sixty-four markers are located on chromosome 10 in a region spreading from ~4.89 to 13.67 Mb. There are five genes in this region that are thought to play a role in controlling stomatal density. All markers within this region exhibit similar allele frequencies, which are correlated with stomatal density and the altitudinal gradient of the stands. We assume this entire region is jointly involved in local adaptation to drought stress. We identified one interesting candidate SNP associated with leaf nitrogen content. Two SNP markers were identified as being significantly associated with $\delta^{13}C$ as measure of intrinsic water use efficiency. Additionally, signals of significant polygenic selection for $\delta^{13}C$ were observed.

Introduction

In forest tree breeding, trees from different populations exposed to specific environmental conditions are sampled and tested in provenance trials or common garden experiments (Neale and Kremer 2011). Adult trees are later tested, because growth and quality trait measurements from juvenile trees are not reliable (Neale and Kremer 2011). In the face of climate change, provenance trials are also used to identify drought tolerant trees with desirable wood quality and growth traits for forest tree breeding programs (Neale and Kremer 2011; Rellstab et al. 2015; Cortés et al. 2020). This step already takes between 20 to 25 years in long-living woody perennials like European beech (*Fagus sylvatica* L.) (Neale and Kremer 2011) and only then the seed and seedling production starts (Neale and Kremer 2011; Cortés et al. 2020). Therefore, it is of great interest in forest tree breeding to determine the genetic basis of adaptive and economic traits and the potential for success of selection on a trait, before we start selection (Cortés et al. 2020; Degen and Müller 2023). Furthermore, identifying candidate gene markers associated with specific traits is desired, as they could be utilized in breeding programs (Neale and Kremer 2011; Cortés et al. 2020; Müller et al. 2024).

Genome-wide association studies (GWAS) aim to identify associations between markers and phenotypic traits of interest (Visscher et al. 2017; Neale and Kremer 2011). GWAS are becoming more popular in forest genetics research (Cumbie et al. 2011; Cappa et al. 2013; Diao et al. 2024; Meger et al. 2024) with increasing availability of reference genomes for tree species (Nystedt et al. 2013; Mishra et al. 2022). A limited number of GWAS have been conducted in European beech (Pfenninger et al. 2020; Müller et al. 2024; Lazic et al. 2024), and even less GWAS used the chromosome-level assembly of European beech (Lazic et al. 2024; Pfenninger et al. 2021).

GWAS return peaks of the "genomic landscape" of adaptation (Rellstab et al. 2015), but according to the "infinitesimal model" of Fisher (1918), quantitative traits are controlled by many small-effect loci. Several methods have been introduced based on Fisher's infinitesimal model (Berg and Coop 2014; Racimo et al. 2018; Zeng et al. 2018, Josephs et al. 2019). Another approach using the "infinitesimal model" is referred to as $\hat{G}$ (Beissinger et al. 2018). $\hat{G}$ follows the model of Wright (1931) that under a selective pressure, alleles with positive effect sizes increase in frequency over generations. In the $\hat{G}$ test, all loci are considered together to retrieve a signal of polygenic selection (Beissinger et al. 2018). $\hat{G}$ is calculated by estimating marker effects with RR-BLUP (Endelmann 2011), an estimation approach from genomic prediction (Beissinger et al. 2018; Mahmoud et al. 2023). But instead of

predicting phenotypes or breeding values, $\hat{G}$ uses the estimated marker effects and combines them with allele frequency differences calculated between different groups that have encountered a selective pressure (Beissinger et al. 2018; Mahmoud et al. 2023). $\hat{G}$ offers an advantage over genomic prediction by providing information on the adaptive or selective pressures that have shaped a trait, and can even indicate the direction of selection. This allows us to determine whether a trait has been subject to negative or positive selection (Beissinger et al. 2018; Mahmoud et al. 2023). $\hat{G}$ can be used to identify signals of polygenic selection when effect sizes are too small to pass the significance threshold of a GWAS or other methods like selection signature mapping or environmental association analysis (Beissinger et al. 2018; Mahmoud et al. 2023). Furthermore, $\hat{G}$ has the advantage over genomic prediction in that its results are easier to interpret (Mahmoud et al. 2023).

$\hat{G}$ does not identify individual markers associated with phenotypes (Beissinger et al. 2018; Mahmoud et al. 2023), but the estimated marker effects can be used for prediction of the performance of individuals (Meuwissen et al. 2001; Endelmann 2011; Alves et al. 2020).

Therefore, $\hat{G}$ is a promising new approach to test for polygenic selection under an environmental selective pressure (Beissinger et al. 2018). Studies in crop species have shown that $\hat{G}$ successfully identifies signals of polygenic selection (Morales et al. 2021; Mahmoud et al. 2023).

Our study comprises 514 individual trees collected in the South-Eastern Carpathian Mountains (Romania) from five European beech stands, along an altitudinal gradient which follows a precipitation gradient. The mean annual precipitation was higher at the higher altitudes; hence the lower stands are more likely to experience periods of limited soil water availability. Several drought stress related and physiological traits were measured. These traits include $\delta^{13}C$ in leaf dry matter as a proxy for water use efficiency (Fotelli et al. 2003; Querejeta et al. 2003) and stomatal density. Stomata, located in the plant leaf epidermis, are important to control gas exchange and transpiration (Gailing et al. 2008; Bertolino et al. 2019; Chen et al. 2022). Stomata characteristics as stomatal density and control are important in managing changing CO_2 levels, temperatures, photosynthetically active radiation and water availability (Gailing et al. 2008; Bertolino et al. 2019; Chen et al. 2022). Furthermore, the leaf nitrogen content and the leaf carbon content were measured, and based on these measurements the C/N ratio was calculated. Leaf nitrogen

content and the C/N ratio are impacted by complex physiological mechanisms (Gessler et al. 2017). At the start of a drought event, lower nitrogen uptake is observed, which later leads indirectly to carbon starvation, which further reduces nitrogen uptake (Gessler et al. 2017). Drought stress can lead to short-term depletion of leaf nitrogen and long-term depletion of leaf carbon (Gessler et al. 2017; Ruehr et al. 2019). The impact on the C/N ratio depends on the nutrient availability before the drought stress event, lower nutrient availability at the beginning of drought stress will result in a higher C/N ratio during drought (Gessler et al. 2017). The correlation of minor allele frequencies (MAF) observed at GWAS candidate loci and traits of interest were further investigated. Nearby annotated genes were referenced based on a literature review to evaluate their putative role in local adaptation.

The findings from this GWAS could help us to prioritize candidate SNPs for drought stress traits in forest breeding programs. If traits are too polygenic to identify candidate SNPs, $\hat{G}$ can provide information on whether traits have a genetic basis and were under polygenic selection. When $\hat{G}$ is successful, the previously estimated marker effects could also be used for genomic prediction. To our knowledge, $\hat{G}$ has not been applied to forest tree species yet. Therefore, we want to test $\hat{G}$ on European beech, a highly outcrossing species, to test if $\hat{G}$ provides insights into genotype-phenotype associations that would remain undiscovered with GWAS.

Material and methods

Genotypes and phenotypes

A data set of 514 individual trees was collected from five beech stands in the South-Eastern Romanian Carpathians along an altitudinal gradient associated with changes in precipitation and temperature (Grigoriadou-Zormpa et al. 2024). The exact geographical locations and more details about the stands are reported by Grigoriadou-Zormpa et al. (2024). The beech stands are referred to as Lempes, Tampa, Solomon, Lupului and Ruia. In Ruia, 110, in Lupului and Lempes 101, in Solomon 102 and in Tampa 99 trees were sampled. Lempes was the lowest elevation site at 550 to 600 m with the lowest mean annual precipitation of 712 mm, followed by Tampa located at 650-700 m with 791 mm, Solomon at 800 to 900 m with 855 mm and Lupului at 1000 to 1150 m with 957 mm mean annual precipitation. Ruia was the highest location at 1300 to 1450 m with the highest mean annual precipitation of 1023 mm. The tree stands are naturally regenerating and were established more than 100 years ago (Grigoriadou-Zormpa et al. 2024).

The phenotypic variables comprised four different traits, which were leaf nitrogen and leaf carbon content, carbon isotope composition $\delta^{13}C$ in leaf dry mass and stomatal density. The leaves of the sun-exposed crowns of the trees were collected in August 2020, 2021 and 2022. Stomatal density was measured by counting the stomata within a fixed area of 0.1938 mm^2 (Yücedağ et al. 2019) on the abaxial side of one dried sun exposed leaf per tree sampled in 2020. The stomata were counted based on a stomatal imprint created with fingernail polish and tape from dried leaves (Yücedağ et al. 2019). Carbon isotope composition ($\delta^{13}C$) in leaf dry mass, leaf nitrogen and carbon content were measured in 2020, 2021 and 2022. To evaluate changes in carbon isotope composition ($\delta^{13}C$) in leaf dry mass, leaf nitrogen and carbon content, homogenized plant material was analyzed by the Centre for Stable Isotope Research and Analysis at the University of Goettingen by elemental analysis combined with isotope ratio mass spectrometer measurements (Werner et al. 1999). Additionally, we calculated the C/N ratio based on the molar content [mmol/g] and molar weights of carbon and nitrogen. Stomatal closure due to reduced soil water availability and altered plant water status reduces the concentration of CO_2 in the stomatal cavities and the concentration of the heavier ^{13}C isotope increases compared to ^{12}C (Fotelli et al. 2003). Consequently, plant water potential can be related to carbon isotope composition ($\delta^{13}C$) (Fotelli et al. 2003). $\delta^{13}C$ is additionally influenced by changing CO_2 levels and photosynthetically active radiation, so this trait shows strong variability across the growing season (Fotelli et al. 2003).

We removed extreme phenotypic outliers. The outliers and distribution of trait measurements are shown in supplemental file S1. Additionally, we examined the correlations of the traits with each other, which are shown in supplemental file S2. The correlations of the traits measured in different years are shown in supplemental file S3. Correlations were analyzed using R version 4.3.1 (R Core Team 2024).

Individual trees were sequenced by using the single primer enrichment technology (SPET) (Scaglione et al. 2019). This sequencing method targets SNPs located within and close to genes and flanking regions of ±2 kb (Scaglione et al. 2019). The target regions were determined by previous whole-genome sequencing of a subset of 96 individual trees by IGATech (IGA Technology Services Srl, Udine, Italy). Reads were aligned to the *F. sylvatica* reference genome version 2 (Mishra et al. 2022) by using BWA-MEM v0.7.17 (Li and Durbin 2009). Only aligned reads with a mapping quality ≥10 were kept and duplicated reads were removed. Variant calling was done along with filtering for minimum read number of an allele in a sample with Freebayes v1.3.6 (Garrison and Marth 2012; Grigoriadou-

Zormpa et al. 2024). Normalization and filtering for read depth > 6 and < 257 were performed using bcftools (Danecek et al. 2011; Grigoriadou-Zormpa et al. 2024). After this, a total of 838,522 SNP markers were kept. The data set was filtered by PLINK for minor allele frequency (MAF) > 0.194% (which results in at least two observations at a marker), missingness of ≤ 10% at each SNP marker and for SNP markers without chromosome annotations. The final data set contained 758,465 SNP markers and 514 individuals used for GWAS. Not for all genotyped individual trees phenotypic measurements were available. For stomatal density, 314 individual trees were phenotyped. Four hundred forty individual trees were phenotyped for the $\delta^{13}C$ measurements in 2020, 457 in 2021 and 454 in 2022. For the leaf nitrogen content measurements, 451 individual trees were phenotyped in 2020, 457 in 2021 and 450 in 2022. For the C/N ratio, 440 individual trees were phenotyped in 2020, 454 in 2021 and 450 in 2022.

GWAS

To test for associations between SNPs and traits, we used a linear model in PLINK v.1.90 (Purcell et al. 2007). All 10 phenotypic variables were tested separately. The first two main components of the principal component analysis (PCA) based on the neutral markers were included as covariates to account for population structure. Neutral SNP markers were identified by Grigoriadou-Zormpa et al. (2024) for population genetic analysis. To identify the neutral SNP markers Grigoriadou-Zormpa et al. (2024) determined intergenic regions by using SnpEff (version 4.3t, Cingolani et al. 2012). Afterwards the filtering based on LD and missingness <10% was done, which resulted in a data set of 154,437 putatively neutral SNPs. To account for environmental effects during sampling in 2020, 2021 and 2022, the first principal coordinate based on precipitation and temperature during the growing season from April until August (supplemental Fig. S4 and supplemental Table S5) in the year of sampling was included (supplemental Fig. S6) in the GWAS for $\delta^{13}C$. This was done because $\delta^{13}C$ shows strong variability across the growing season (Fotelli et al. 2003). Temperature and precipitation weather data were downloaded from Copernicus Climate Data Store (CDS 2024) for 2020, 2021 and 2022 (available in supplemental Fig. S4, Table S5 and Fig. S6). The principal coordinate analysis (PCoA) was performed on monthly temperature and precipitation measured from April until August for the years separately using the *cmdscale()* R function in the R version 4.3.1 (R Core Team 2025).

The significance thresholds were derived based on a permutation test using the same set of SNP markers (Purcell et al. 2007). The permutation test was conducted in the adaptive permutation mode, which can terminate before

completing all possible iterations, when there are no changes anymore (Purcell et al. 2007). This adaptive permutation mode was run with ten minimal and 1,000,000 maximal permutations. The level at which no markers were identified as significant in our permutation test, was declared the significance threshold, which in our case was $p \leq 0.000001$, which was the same for all traits. QQ plots are shown in supplemental file S7.

Gene variant annotation

SnpEff (version 4.3t, Cingolani et al. 2012) was used for gene variant annotations. An annotation data base was built based on the *F. sylvatica* reference genome (Mishra et al. 2022), which was then also used in this study for the gene variant annotation. SnpEff uses this data base and the VCF file from this study to annotate gene variants, their effects and predicted impacts (Cingolani et al. 2012). The effects comprise missense variants, variants causing stop codons, frameshift variants, splice region variants and several more (Cingolani et al. 2012). A missense variant causes a codon change that produces a different amino acid (Cingolani et al. 2012).

Ghat

Groups of stands growing under different environmental conditions were treated as separate populations to implement $\hat{G}$ (Fig. 1). The groups were formed on the basis of altitude and the amount of precipitation (Fig. 1). Group 1 included the stands Ruia and Solomon which were located at altitudes > 800 m and with a mean annual precipitation > 850 mm (Fig. 1). Group 2 included the stands Tampa and Lempes, which were all located at altitudes < 800 m and with a mean annual precipitation < 800 mm (Fig. 1). Lupului was removed from the $\hat{G}$ analysis, because of the age of the stand in comparison to the other stands (Fig. 1). The marker effect estimates were calculated, by using the rrBLUP R package (Endelman 2011).

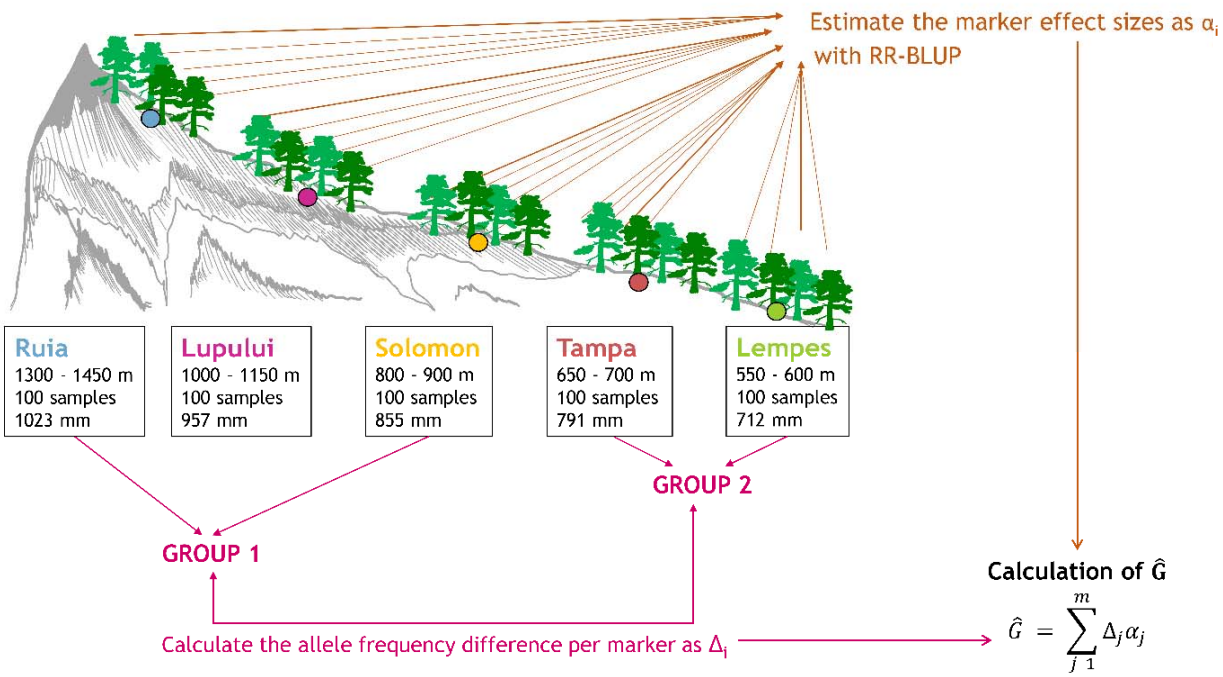


Fig. 1: Simplified scheme of the location of the beech stands along the altitudinal gradient and the calculation of by evaluating the marker effect sizes and allele frequency differences between groups of different environmental conditions across the entire genome.

Filtering for minor allele frequency (MAF) and missingness <0.05 was done before missing observations were imputed based on the mean allele frequency of the stands observed at each marker in R version 4.3.1 (R Core Team 2024). When was calculated for $\delta^{13}C$, the weather data of the different years of sampling (see supplemental Figure S4 and Table S5) was included in the model to estimate the marker effects. was calculated with the Ghat R package (Mahmoud et al. 2023). The number of permutations was set to 1000. The number of effective markers was first calculated based on a function called "ld_decay", which is provided by the Ghat package (Mahmoud et al. 2023). To evaluate its impact on , we additionally tested 100 randomly generated traits. These 100 random traits were generated by simulating a normal distribution with a mean of 25.93 and a standard deviation of 3.67 (mean and standard deviation of the trait leaf nitrogen content), then the traits were randomly associated to the individual trees. In the test runs we also tested if these random traits are identified as being under significant selection. When the random traits were identified as being under significant selection, they were considered to be false positive observations. Based on the 100 random traits, a rate of false positive observations was calculated and used to evaluate the number of effective markers. Later, in the actual test, we tested the four different traits stomatal density, $\delta^{13}C$, leaf nitrogen content and C/N ratio to identify signals of polygenic selection.

Cross-validation of marker effect sizes

To evaluate the marker effect estimates used in the $\hat{G}$ test, we conducted a cross-validation. For the cross-validation, the data set was randomly partitioned into five sets to be used in five-fold cross-validation with 100 replications (Endelmann 2011). In the cross-validation, the phenotypic values per trait of the testing set were predicted based on marker effect estimates which were calculated based on the training set (Endelmann 2011). The marker effect estimates were calculated with rrBLUP, using the same method as for the $\hat{G}$ test (Endelmann 2011). The predictive ability was calculated as correlation of predicted trait estimates and observed trait measurements for the different traits of the testing set (Endelmann 2011). Additionally, another cross-validation was conducted in which one stand was left out and the remaining stands were used for prediction. With this cross-validation we could assess how well the stands can predict a new stand.

Calculation of linkage disequilibrium (LD) decay

We created a pairwise LD heat map with the R package LDheatmap for the regions which were associated with the traits and putatively under selection (Shin et al. 2006).

Results

GWAS revealed a peak of markers on chromosome 10 strongly associated with stomatal density and to a smaller extent with $\delta^{13}C$

Stomatal density was significantly associated with 121 markers, of which 101 are located on chromosome 10 (Fig. 2). The remaining markers are located on chromosomes 1, 4, 5, 6, 8 and 9, with one to five markers on each chromosome (Fig. 2). The 101 markers on chromosome 10 are located between ~4.89 to ~13.67 Mb, which appear as one wide peak (Fig. 2).

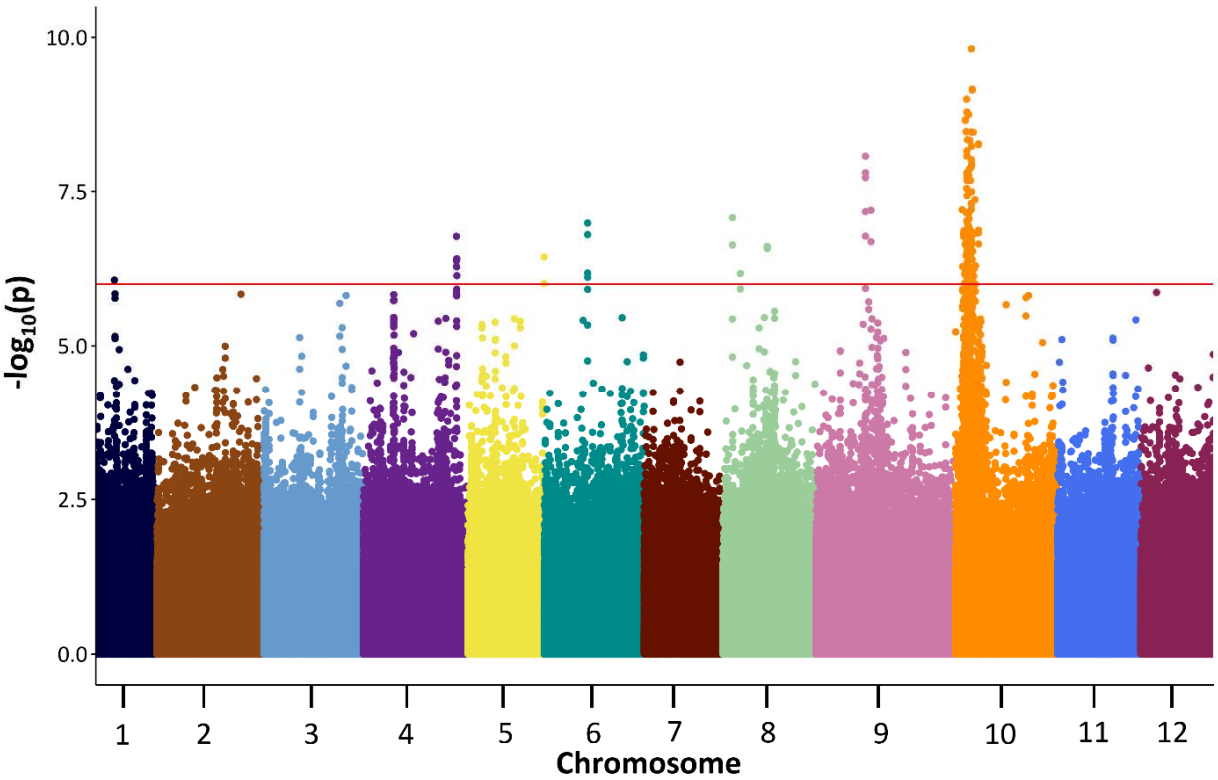


Fig. 2: Manhattan plot for stomatal density with premutation derived significance thresholds of p ≤0.000001 (red horizontal line) with p-values plotted as $-\log_{10}(p)$ against the position in bp.

A similar peak on chromosome 10 appears in the GWAS for $\delta^{13}C$ measured in 2020 and 2022 (Fig. 3). But this peak is smaller and contains almost no significant SNP markers (Fig. 4.3). Only for $\delta^{13}C$ measured in 2020, two SNP markers exceeded the permutation derived significance levels of p ≤0.000001 (Fig. 3). No significant SNPs were associated with $\delta^{13}C$ measured in 2021 and 2022 (Fig. 3).

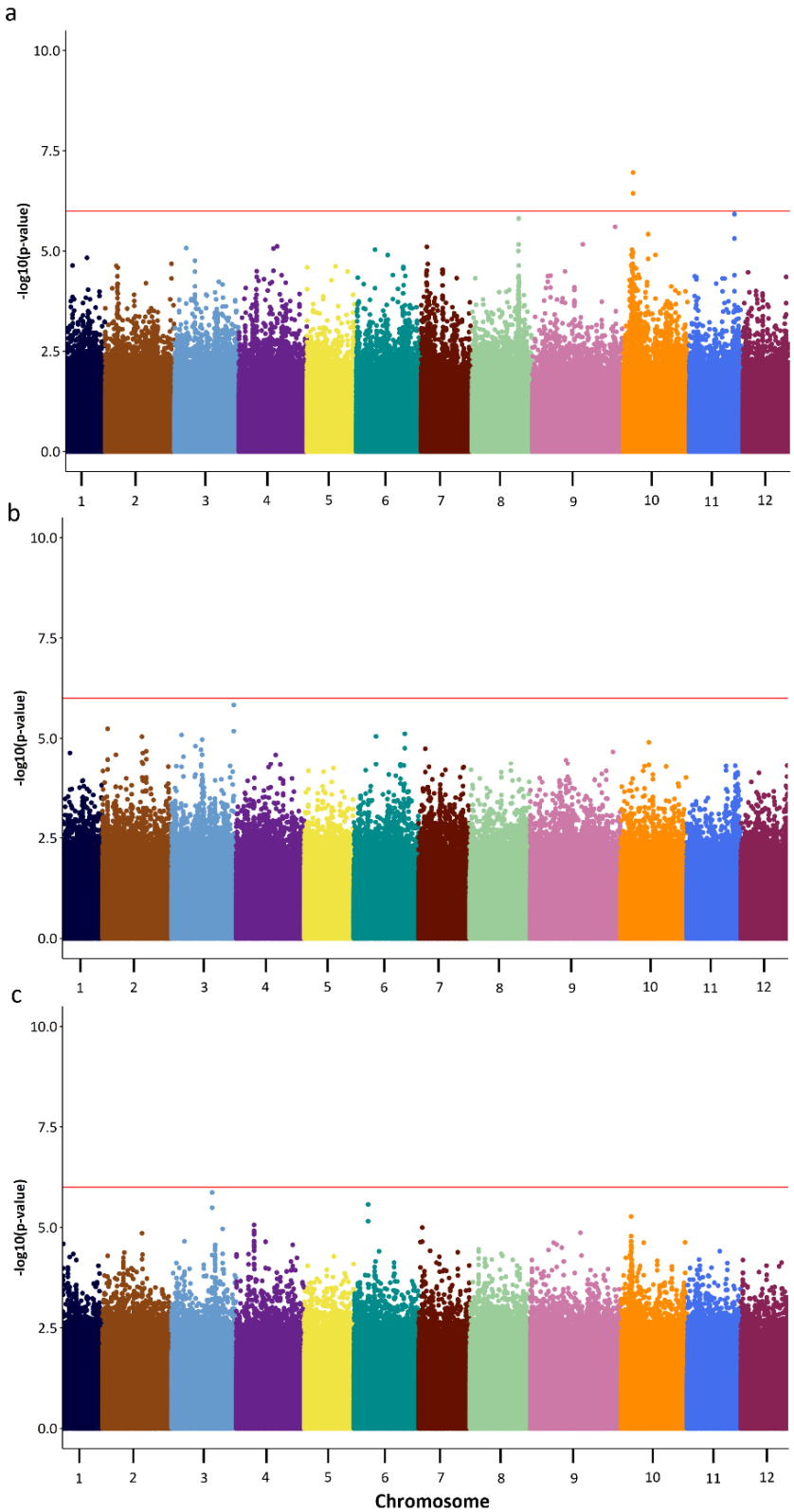


Fig. 3: Manhattan plot for $\delta^{13}C$ measured in 2020 (**a**), 2021 (**b**), and 2022 (**c**) with permutation derived significance thresholds of $p \leq 0.000001$ (red horizontal line) with p-values plotted as -log10(p) against the position in bp.

Table 1 presents the most interesting GWAS results for stomatal density, focusing on candidate SNPs within 20 kb of gene annotations. The whole list of significant SNPs for stomatal density and $\delta^{13}C$ measured in 2020 can be found in supplemental Table S8. The SNP markers significantly associated with $\delta^{13}C$ measured in 2020 were not located within 20 kb of impactful parts of coding regions of genes (supplemental Table S8).

Table 1: List of candidate SNPs with gene annotations associated with stomatal density with permutation derived significance levels of p ≤0.000001 and their annotated gene variants, the underlying genes with their locations, functions and predicted impact (HIGH, MOD (moderate) and LOW) based on SnpEff (Cingolani et al. 2012). Candidate genes involved in stomatal regulation, chloroplast biogenesis, and chloroplast metabolism are highlighted in bold face.

Chr	Position of SNP	Distance to gene variant in kb	Gene variant		Impact	Gene
8	3998414	15.19 to 15.41	Bhaga_8.g506	8 missense variants	MOD	VVA35186.1: uncharacterized LOC117631844 (NCBI 2025)
	15485128	~-6.91 to -6.99	Bhaga_8.g1863	1 missense variant	MOD	**XP_030937567.1:** ABC transporter I family member 10 (NCBI 2025)
	15485213	0.98				
9	13529604	~-12.782	Bhaga_9.g1680	1 missense variant	MOD	XP_030932484.1: transcription factor bHLH117 (NCBI 2025)
	13556977	~1.942 to 3.746	Bhaga_9.g1682	13 missense variants	MOD	XP_030928910.1: probable terpene synthase 9 (NCBI 2025)
	13557019	~1.94 to 3.704 kb	Bhaga_9.g1682			
10	4898376	located within the coding region	Bhaga_10.g591	Synonymous gene variant		XP_030959197.1: proteasome adapter and scaffold protein ECM29 (NCBI 2025)
		~4.8		3 missense variants	MOD	
	6314154	~-18.5	Bhaga_10.g752	1 missense variant	MOD	XP_030939072.1: histone deacetylase HDT1-like (NCBI 2025)
	6314154	~9.4	Bhaga_10.g753	3 missense variants	MOD	VVA25679.1: uncharacterized LOC117631118 (NCBI 2025)
	7116359	~-3.0	Bhaga_10.g852	1 missense variant	MOD	VVA33914.1: uncharacterized LOC117632143 (NCBI, 2025)
	7116359	~-9.6	Bhaga_10.g854	1 frameshift variant	HIGH	XP_030962073.1: germin-like protein 9-3 (NCBI 2025)
	7116359	~-9.7	Bhaga_10.g854	7 missense variants	MOD	
			Bhaga_10.g854	disruptive inframe insertion		
	7198804	~13.6 to 13.8	Bhaga_10.g867	7 missense variants	MOD	XP_030956796.1: capsanthin/capsorubin synthase, chromoplastic-like (NCBI 2025)
	7211779	-18.4	Bhaga_10.g873	Variant hits 5'UTR region with premature start codon gain variant	LOW	XP_030956765.1: mRNA-decapping enzyme-like protein (NCBI, 2025)
	7555433	3.1	Bhaga_10.g917	stop codon	HIGH	**XP_030964327.1:** aluminum-activated malate transporter 2-like isoform X1 (NCBI 2025)
	7555433	~3.2 to 3.3		9 missense variants	MOD	

7785927	~15.9 to 16.1	Bhaga_10.g957	30 missense variants	MOD	**XP_030940718.1:** transcription factor MYB35-like isoform X2 (NCBI 2025)
7785978					
7787468					
7820271					
7820389					
11903957	14.3	Bhaga_10.g1460	3 missense variants	MOD	**XP_030939393.1:** ATP-dependent Clp protease adapter protein *ClpS1*, chloroplastic-like (NCBI, 2025)
11903957	~-11.5 to -11.7	Bhaga_10.g1462	5 missense variants	MOD	**XP_030938815.1:** protein ASPARTIC PROTEASE IN GUARD CELL 1 (NCBI, 2025)
12547109	located within the coding region	Bhaga_10.g1543	unknown (missing information on the reading frame)	NA	**XP_030937177.1:** glycosyltransferase BC10-like (NCBI, 2025
	~-0.3 to -0.4		3 missense variants	MOD	
12667233	~-17.5	Bhaga_10.g1562	2 splice region variants and two intron variants	LOW	**XP_030939348.1:** kinesin-like protein NACK1 (NCBI, 2025)
13435289	located within the coding region	Bhaga_10.g1651	Synonymous gene variant		**XP_023902123.1:** protein BONZAI1 (NCBI, 2025)
	~4.8		1 splice region variant	LOW	
	~-14.6	Bhaga_10.g1652	1 missense variant	MOD	VVA19309.1: uncharacterized LOC117629398 (NCBI, 2025)
13677756	~-11.8	Bhaga_10.g1674	22 missense variants	MOD	XP_034215671.1: poly(A) polymerase I-like isoform X1 (NCBI, 2025)
13677816					

At the marker on chromosome 10 at ~13.435 Mb, the MAF increases with increasing stomatal density which follows the altitudinal gradient associated with precipitation and temperature (Fig. 4). For all other markers on chromosome 10 from ~4.89 to 13.67 Mb, we observed the same trend (see supplemental Table S9 and Fig. S10). In all cases, the minor allele frequency (MAF) is correlated with stomatal density (see supplemental Table S9 and Fig. S10).

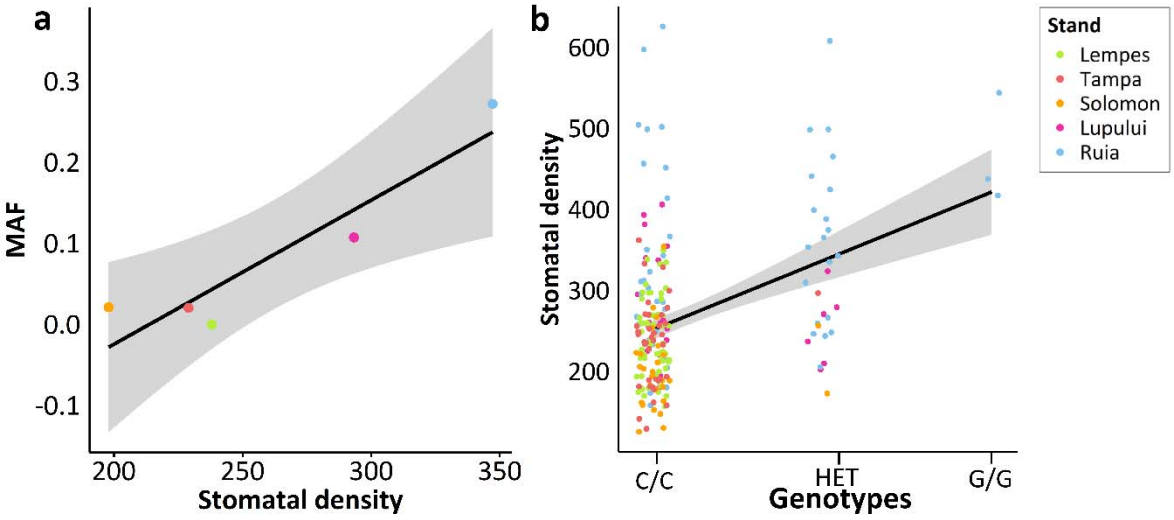


Fig. 4: a) Correlation between stomatal density and the minor allele frequency (MAF) of the significant marker (p ≤0.000001) on chromosome 10 at ~ 13.435 Mb with a correlation coefficient of 0.932 (p-value = 0.021) and **b**) Linear regression of marker genotypes against stomatal density with a coefficient of determination R^2 of 0.1483 (p = 0.0001258). This marker was annotated with gene variant *Bhaga_10.g1651*, with the underlying gene *XP_023902123.1*, also described as *BONZAI1*.

Associations of SNP markers with leaf nitrogen and carbon traits differed across years of sampling

Figure 5 displays the Manhattan plots for leaf nitrogen content and C/N ratio measured in 2020, 2021 and 2022, which were the other traits with significant associations. The SNP markers associated with the same trait measured in different years did not overlap (Fig. 5).

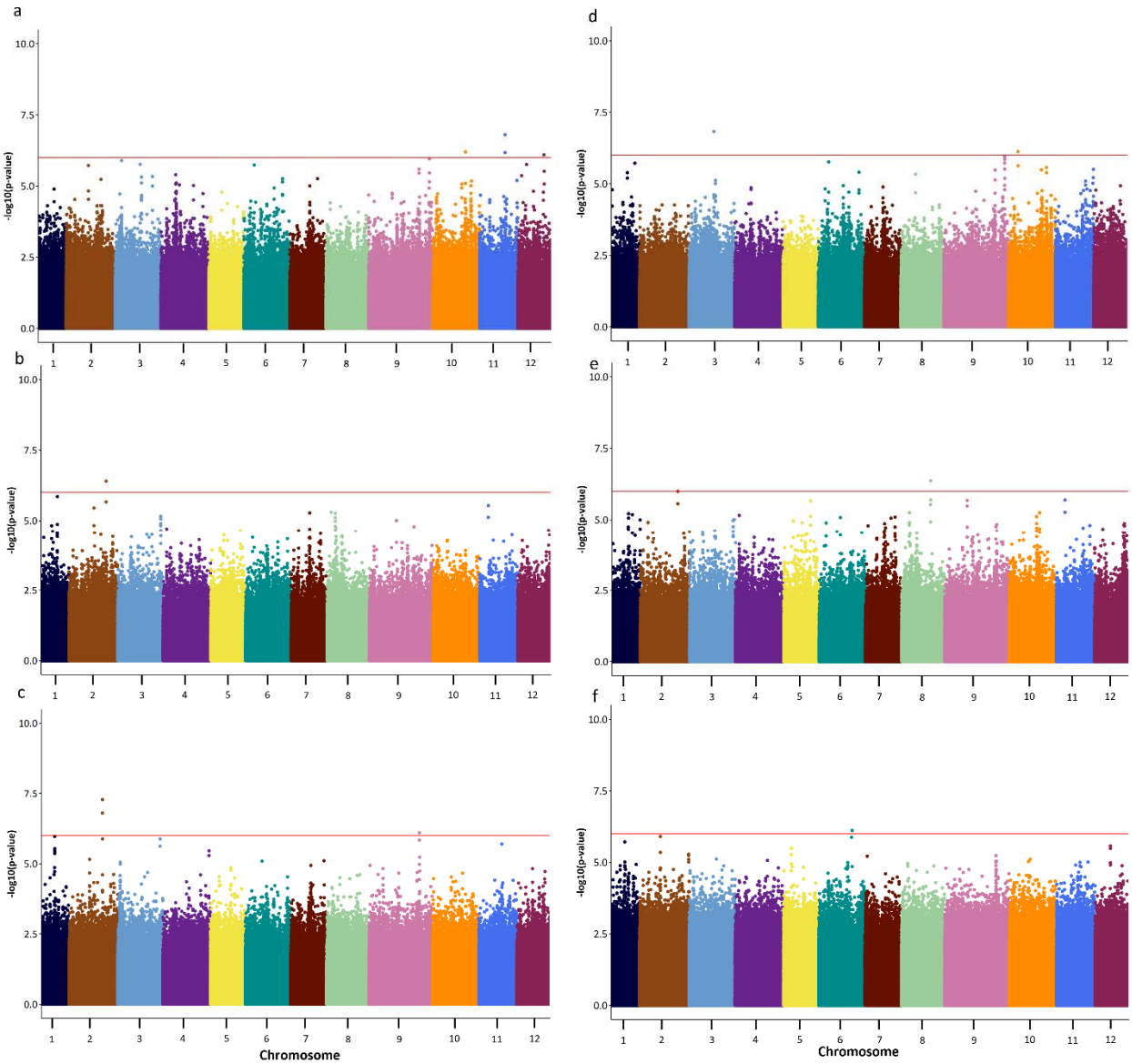


Fig. 5: Manhattan plot for leaf nitrogen content measured in 2020 (**a**), 2021 (**b**), and 2022 (**c**) and C/N ratio measured in 2020 (**d**), 2021 (**e**), and 2022 (**f**) with permutation derived significance thresholds of p ≤0.000001 (red horizontal line) with p-values plotted as $-\log_{10}(p)$ against the position in bp.

Table 2 shows the GWAS results for leaf nitrogen and carbon traits, which include only candidate SNPs of genes, which were < 20 kb away from the coding region of genes. These candidate SNPs were associated with leaf nitrogen content and C/N ratio measured in 2020, 2021, 2022 (Table 2).

Table 2: List of candidate SNPs associated with the other drought-stress related traits NC (leaf nitrogen content) and CNR (C/N ratio) with permutation derived significance levels of p ≤0.000001 and their annotated gene variants, the underlying genes with their locations, functions and predicted impact (HIGH, MOD (moderate) and LOW) based on SnpEff (Cingolani et al. 2012).

Trait	Chr	Position of SNP	Distance to gene variant in kb	Annotated gene variants		Impact	Gene
NC (2022)	2	30263171	NA	NA		NA	NA
NC (2022)	2	30263199	-3.26	Bhaga_2.g3283	1 missense variant and a conservative inframe deletion	MOD	XP_023907498.1
			~-3.25 to 3.4		5 missense variants	MOD	
NC (2021)	2	31803661	~1.8 to 4.983	Bhaga_2.g3456	4 missense variants	MOD	POO01414.1
			~-4.89 to -13.99	Bhaga_2.g3457	5 missense variants	MOD	**XP_030960280.1:** Abscisic-aldehyde oxidase-like
CNR (2020)	3	29431182	~7.995	Bhaga_3.g3375	1 stop gain variant	HIGH	NA
			~8.19 to 8.2		2 frameshift variants	HIGH	
			~7.99 to 8.182		5 missense variants	MOD	
CNR (2022)	6	27342063	-3.24	Bhaga_6.g3025	synonymous variant		XP_031278606.1
CNR (2021)	8	21323750	17.68	Bhaga_8.g2539	1 splice region variant	LOW	XP_023876917.1
			~6.1 to 6.13	Bhaga_8.g2541	3 missense variants	MOD	RVW32758.1
				Bhaga_8.g2541			
				Bhaga_8.g2541			
NC (2022)	9	26070622	-5.84	Bhaga_9.g3146	1 missense variant	MOD	RVW57531.1
CNR (2020)	10	13027242	NA	NA	NA	NA	NA
NC (2020)	10	39245380	NA	NA	NA	NA	NA
NC (2020)	11	26539169	~8.56 to 8.57	Bhaga_11.g3150	2 missense variants	MOD	XP_030948806.1 : protein NIM1-INTERACTING 1-like (NCBI, 2025)
NC (2020)	12	39062439	NA	NA	NA	NA	NA

For C/N ratio and leaf nitrogen content, we do not observe a consistent trend for the significant markers (see supplemental Table S10, Figure S12 and S13). Supplemental Figure S11 shows the correlations for MAF and leaf nitrogen content in 2021 at the marker on chromosome 2 at ~31.804 Mb, which was annotated with Bhaga_2.g3457 (*abscisic-aldehyde oxidase-like).*

Ghat indicates signals of positive polygenic selection for leaf $\delta^{13}C$

To evaluate the calibration of $\hat{G}$, we tested 100 randomly generated traits. The calculated number of effective markers estimated based on LD decay over the entire genome was 838,522, which corresponds to the number of total markers. This resulted in 90 out of our randomly generated traits appearing as under selection by the uncalibrated $\hat{G}$ test. We did a grid search for effective marker number (ranging from 1,000 to 800,000). With 7,000 effective markers, we observed a rate of false positive observations <10% (see supplemental file S14). Results from cross-validation of predictions calculated with marker effect estimates based on 100 replications are available in supplemental file S15.

We calculated $\hat{G}$ between the high precipitation (Ruia and Solomon) and low precipitation group (Tampa and Lempes) and observed positive selection with a correlation coefficient between allele effect estimates and allele frequency change of 0.1017 (p ~0) and 0.0962 (p ~0) for $\delta^{13}C$ in 2020 and 2021 (see supplemental Table S16). Based on this

observation, we can deduce that polygenic selection for $\delta^{13}C$ decreased with increasing precipitation along the altitudinal gradient (see also supplemental Table S1). $\hat{G}$ calculated based on $\delta^{13}C$ measured in 2022 was not significant (supplemental Table S16). Soil water availability and plant water status can be related to $\delta^{13}C$, because an increase in $\delta^{13}C$ describes the reduced concentration of CO_2 in the stomatal cavities under closed stomata to improve the plant water potential (Fotelli et al. 2003). Under these circumstances, water use efficiency is usually increased (Fotelli et al. 2003).

Discussion

Annotations of genes harboring SNP markers significantly associated with stomatal density underpin a functional role in stomatal patterning and regulation of leaf water status

Stomatal density was significantly associated with 121 different markers, of which 101 are located on chromosome 10. In total, 26 of these markers, are located between position ~4.89 to ~13.67 Mb on chromosome 10 and appear as one wide peak. Across this region we observed increased linkage disequilibrium in the stands Solomon, Tampa and Lupului between R^2 of 0.5 to 0.6. A total of 16 genes was identified within this peak, five of which are reported to have a direct influence on stomata regulation, chloroplast metabolism, chloroplast biogenesis or plastid maintenance

(Gou et al. 2015; Nishimura et al. 2015; Shen et al. 2018; Dabravolski and Isayenkov 2023; Wang et al. 2024; Frangedakis et al. 2024) in *Arabidopsis thaliana* L. Overall, we found gene annotations for 21 genes of candidate SNPs associated with stomatal density on four chromosomes, but most interesting genes were annotated for gene variants located on chromosome 10.

Another gene with a potential influence on stomata was annotated for two markers on chromosome 8 at ~15.485 Mb. This gene is called *ABC transporter I family member 10* (NCBI 2024). The two markers on chromosome 8 at ~15.485 Mb are 0.98 kb upstream from a missense variant in the coding region of this gene variant. *ABC transporter I family member 10* belongs to the family of ATP-binding cassette (ABC) transporter lipids (Voith van Voithenberg et al. 2019). In *Arabidopsis thaliana*, *ABC transporter I family member 10* was observed to be responsible for ABC protein generation (Voith van Voithenberg et al. 2019). Loss of this gene led to extremely dwarfed albino plants with impaired chloroplasts (Voith van Voithenberg et al. 2019). Strong LD was observed between the two markers on chromosome 8 at ~15.485 Mb and the coding region of *Bhaga_8.g1863 (ABC transporter I family member 10)* (supplemental Fig. S17).

The genes, located on chromosome 10, with a direct influence on stomatal characteristics comprise *aluminum-activated malate transporter 2-like* (Dabravolski and Isayenkov 2023), *transcription factor MYB35-like* (Frangedakis et al. 2024), *ATP-dependent Clp protease adapter protein ClpS1* (Nishimura et al. 2015), *aspartic protease in guard cell 1* (Shen et al. 2018; Wang et al. 2024) and *BONZAI1* (Gou et al. 2015).

The gene variant *Bhaga_10.g917* corresponds to the underlying gene *aluminum-activated malate transporter 2-like (ALMT2)* (NCBI 2024). *Aluminum-activated malate transporter 2-like (ALMT2)* belongs to the aluminum-activated malate transporter (ALMT) gene family which is involved in tolerance to Al^{3+} and regulation of opening and closing of stomata in plants (Dabravolski and Isayenkov 2023). A stop codon of this gene is located ~3.1 kb upstream from the marker found on chromosome 10 at ~7.56 Mb that was associated with stomatal density (see supplemental Fig. S18). This stop codon occurs in the first third of the coding region of the gene variant (see supplemental Fig. S18). Furthermore, nine missense variants of the gene coding region are located ~3.2 to 3.3 kb upstream from the marker. According to the pairwise LD (linkage disequilibrium) heat map analysis, we observed strong LD around the marker and this gene (see supplemental Fig. S18). All nine missense variants caused a codon change that resulted in a

different amino acid sequence (see supplemental Table S19). At six missense variants a hydrophilic amino acid was replaced by a hydrophobic amino acid (see supplemental Table S19). The change from a hydrophilic amino acid to a hydrophobic amino acid can increase the stability of the sequenced proteins in vivo (Lindman et al. 2010). At two out of nine missense variants the amino acids phenylalanine and serine were replaced by leucine (see supplemental Table S19).

The gene variant *Bhaga_10.g957* corresponds to the underlying gene *transcription factor MYB35-like* (NCBI 2024). MYB related transcription factors like *transcription factor MYB35-like* control chloroplast biogenesis and target genes for carbon fixation and light harvesting (Frangedakis et al. 2024). Mutants of MYB-related genes appear bleached and contain small chloroplasts (Frangedakis et al. 2024). We found five markers on chromosome 10 at ~7.78 to ~7.82 Mb which were ~16 kb upstream from 30 missense variants located within the coding region of *transcription factor MYB35-like* with a moderate impact on its expression (see also supplemental Fig. S20). Our pairwise LD heat map, available as supplemental Figure S20, shows that the significant marker is in very strong and almost complete LD (~1) with the coding region of this gene. No change in the polar bonds was detected in 25 missense variants (see supplemental Table S19). At three missense variants a hydrophilic amino acid was replaced by a hydrophobic amino acid (see supplemental Table S19). Glutamic acid was replaced by lysin, alanine was replaced by valine, and arginine was replaced by glutamine (supplemental Table S19). In the remaining two missense variants, one hydrophobic amino acid was replaced by a hydrophilic amino acid (see supplemental Table S19).

For the gene variant *Bhaga_10.g1460*, the annotated gene *ATP-dependent Clp protease adapter protein ClpS1* also referred to as *ClpS1* was found (NCBI 2024). *ClpS1* forms together with *ClpF* an important unit in the chloroplasts which serves as an adaptor and is responsible for substrate recognition and delivery (Nishimura et al. 2015). Our identified marker on chromosome 10 at ~ 11.9 Mb is located ~14.3 kb downstream from three missense variants of the coding region of the gene variant *Bhaga_10.g1460* (*ClpS1)* (see also supplemental Fig. S21). At two missense variants of *Bhaga_10.g1460* (*ClpS1)*, a hydrophilic amino acid was replaced by a hydrophobic amino acid (see supplemental Table S19).

The marker was located in the intergenic region between *Bhaga_10.g1460* (*ClpS1)* and *Bhaga_10.g1462* (*ASPG1*), but the pairwise LD is only increased in some parts along this region (see also supplemental Fig. S21). The gene variant *Bhaga_10.g1462* was annotated as the gene *aspartic protease in guard cell 1*, also known as *ASPG1* (NCBI 2024). Stomata pores are surrounded by guard cells (Negi et al. 2013). *ASPG1* was found to be overexpressed under drought stress in *Arabidopsis* (Yao et al. 2012). The overexpression of the *ASPG1* gene increases the abscisic acid (ABA) sensitivity in stomatal guard cells and leads to reduced water loss in transgenic *Arabidopsis* plants (Yao et al. 2012). Shen et al. (2018) also observed that this gene plays a role in seed dormancy, seed longevity and seed germination in *Arabidopsis*. In general, aspartic proteases are involved in various biological processes like chloroplast metabolism, biotic and abiotic stress responses and reproductive development (Wang et al. 2024). Our identified marker on chromosome 10 at ~11.9 Mb is also located ~12 kb downstream from five missense variants of *Bhaga_10.g1462* (*ASPG1*) (see also supplemental Fig. S21). At three missense variants of *Bhaga_10.g1462* (*ASPG1*), a hydrophilic amino acid was replaced by a hydrophobic amino acid (see supplemental Table S19).

For the gene variant *Bhaga_10.g1651*, the gene *BONZAI1* was annotated (NCBI 2024). BONZAI genes control global osmotic stress responses, ABA accumulation and stomata regulation in plants (Chen et al. 2020). *BONZAI1* regulates Ca^{2+} signaling under osmotic stress in rice (Chen et al. 2020). In *Arabidopsis*, *BONZAI1* was found to also promote stomatal closure and ABA accumulation (Gou et al. 2015). ABA plays an important role in stomatal closure (Chen et al. 2022). The significant marker on chromosome 10 at ~13.44 Mb is located within the coding region of *Bhaga_10.g1651* (*BONZAI1*) (synonymous gene variant) and ~4.8 kb downstream from this marker is a splice region variant of this gene (see supplemental Fig. S22). The gene variant *Bhaga_10.g1652* is located ~14 kb downstream from this marker, but no gene description was found for this gene variant (see supplemental Fig. S22). One missense variant was found in the coding region of *Bhaga_10.g1652*, which is ~14.6 kb downstream of the significant marker on chromosome 10 at ~13.44 Mb (supplemental Fig. S22).

The remaining genes are involved in disease resistance, degradation and regulation processes. A potential role in stomatal density could not be directly inferred from the literature and further studies are needed. These comprise *histone deacetylase HDT1-like*, *germin-like protein 9-3*, *capsanthin synthase chromoplastic-like*, *mRNA-decapping enzyme-like protein*, *glycosyltransferase BC10-like*, *kinesin-like protein NACK1* and *poly(A) polymerase I-like* (NCBI 2024). Their description is available in supplemental Table S24.

We assume that the genes *histone deacetylase HDT1-like* (*Bhaga_10.g752*) and *glycosyltransferase BC10-like* (*Bhaga_10.g1543*) could also be important for stomatal density, since they are involved in stem vascular development and cellulose biosynthesis. Chhetri et al. (2020) detected genes encoding glycosyltransferases in a GWAS for stomatal density. Our significant marker on chromosome 10 at ~12.5471 Mb is located within the coding region of *Bhaga_10.g1543 (glycosyltransferase BC10-like)*, but at this position the impact on the coding of the gene is unknown (missing information on the reading frame) (see supplemental Fig. S23). We observed very strong LD (~1) across the entire coding region of this gene (see supplemental Fig. S23).

In contrast, *Capsanthin synthase chromoplastic-like* does not seem to be important for stomatal density. The importance of the other genes for stomatal density is also difficult to deduce. It may be possible that some of these genes are hitchhiking, because strong joint selection is acting across the entire region on chromosome 10 from ~4.89 to ~13.67 Mb.

Significant correlations of stomatal density with minor allele frequency and environmental parameters indicate a contribution to local adaptation

All markers located on chromosome 10 between ~4.89 to ~13.67 Mb show a positive correlation between stomatal density and MAF (Fig. 4, supplemental file S9 and S10). In this study, we observed increasing stomatal density with increasing altitude and precipitation and decreasing temperature. Hamanishi et al. (2012) observed in *Populus* that lower stomatal density restricts the number of sites for water loss, thereby reducing transpiration. Li et al. (2023) found that under drought stress, *Populus* trees exhibit lower stomatal densities but increased stomatal length and width and improved stomatal control mechanisms. In this study we observed a correlation of stomatal density with water use efficiency measured as $\delta^{13}C$ measured in 2020 at 0.23 ($p \leq 0.005$), in 2021 at 0.18 ($p \leq 0.005$) and in 2022 at 0.18 ($p \leq 0.001$) (supplemental Fig. S2). Overall, we found slightly more genes involved in stomatal control and regulation than in stomatal development. Stomatal density and stomatal conductance are usually positively correlated, in this study stomatal conductance was not measured (Franks et al. 2009).

We observed very similar correlations between stomatal density and MAF across all markers on chromosome 10 between ~4.89 to ~13.67 Mb (Fig. 4, supplemental file S9 and S10). This indicates joint selection for this entire

region, in which at least five genes with a direct influence on stomata are located. The correlation of MAF and stomatal density is not increasing in full agreement with the position of the stands along the altitudinal gradient (Fig. 4, supplemental file S9 and S10). Solomon and Tampa have slightly lower stomatal density than Lempes. This might be explained by the steeper slopes in Solomon and Tampa than in Lempes, which may have led to more water being retained over the years in Solomon and Tampa in comparison to Lempes.

Figure 6 displays the observed haplotypes across the entire region on chromosome 10 between ~4.89 to ~13.67 Mb and the pairwise LD heat map within the different stands. Observed haplotypes across the entire region differ especially between Ruia and Lupului, Tampa and Lempes (Fig. 6A). Pairwise LD is increased in Lupului, Solomon and Tampa in comparison to Ruia (Fig. 6B).

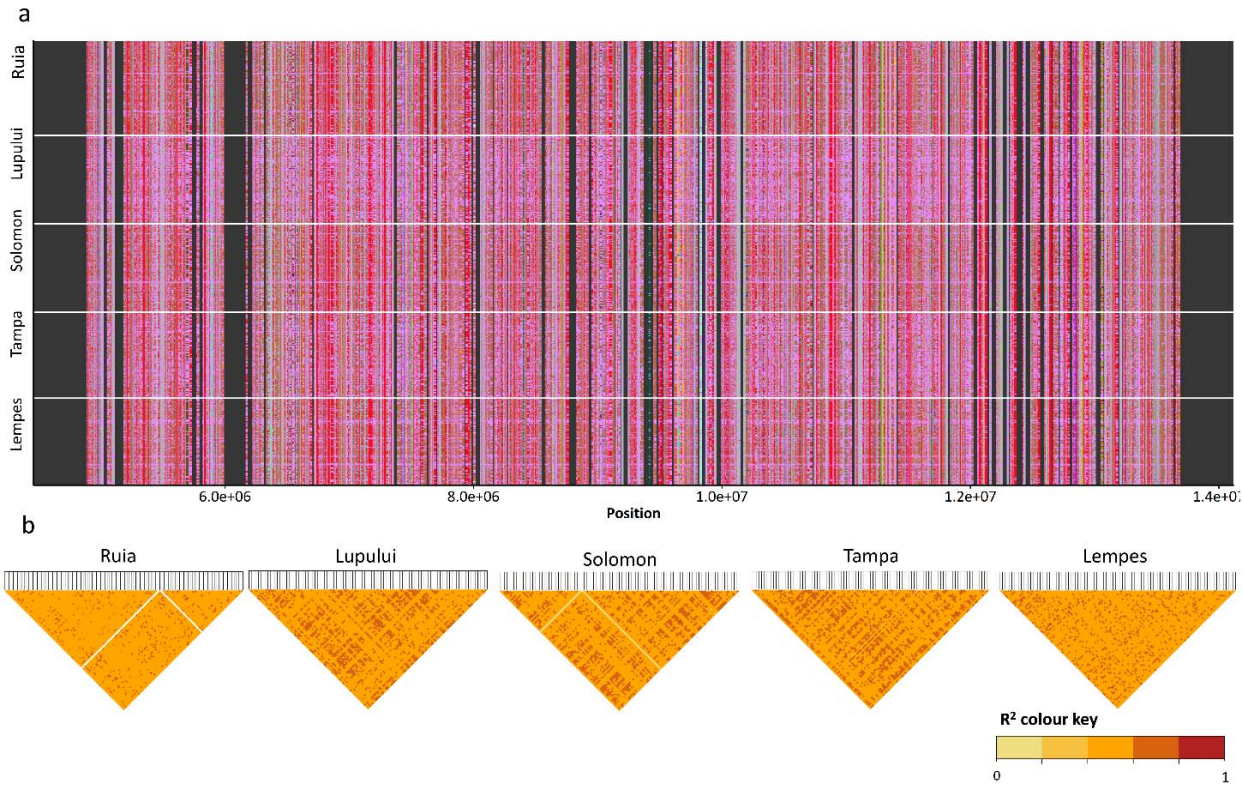


Fig. 6: Region on chromosome 10 from ~4.89 to ~13.67 Mb and the observed haplotypes (**a**) and pairwise LD heat map for the different subpopulations (**b**).

Further evidence supporting that this region on chromosome 10 is involved in local adaptation, is provided in Figure 7. F_{ST}-based selection signature mapping revealed a peak on chromosome 10 from ~4.512 to ~47.742 Mb (Fig. 7).

This peak is partly overlapping with the peak containing marker associations from the GWAS for stomatal density from ~4.89 to ~13.67 Mb on chromosome 10.

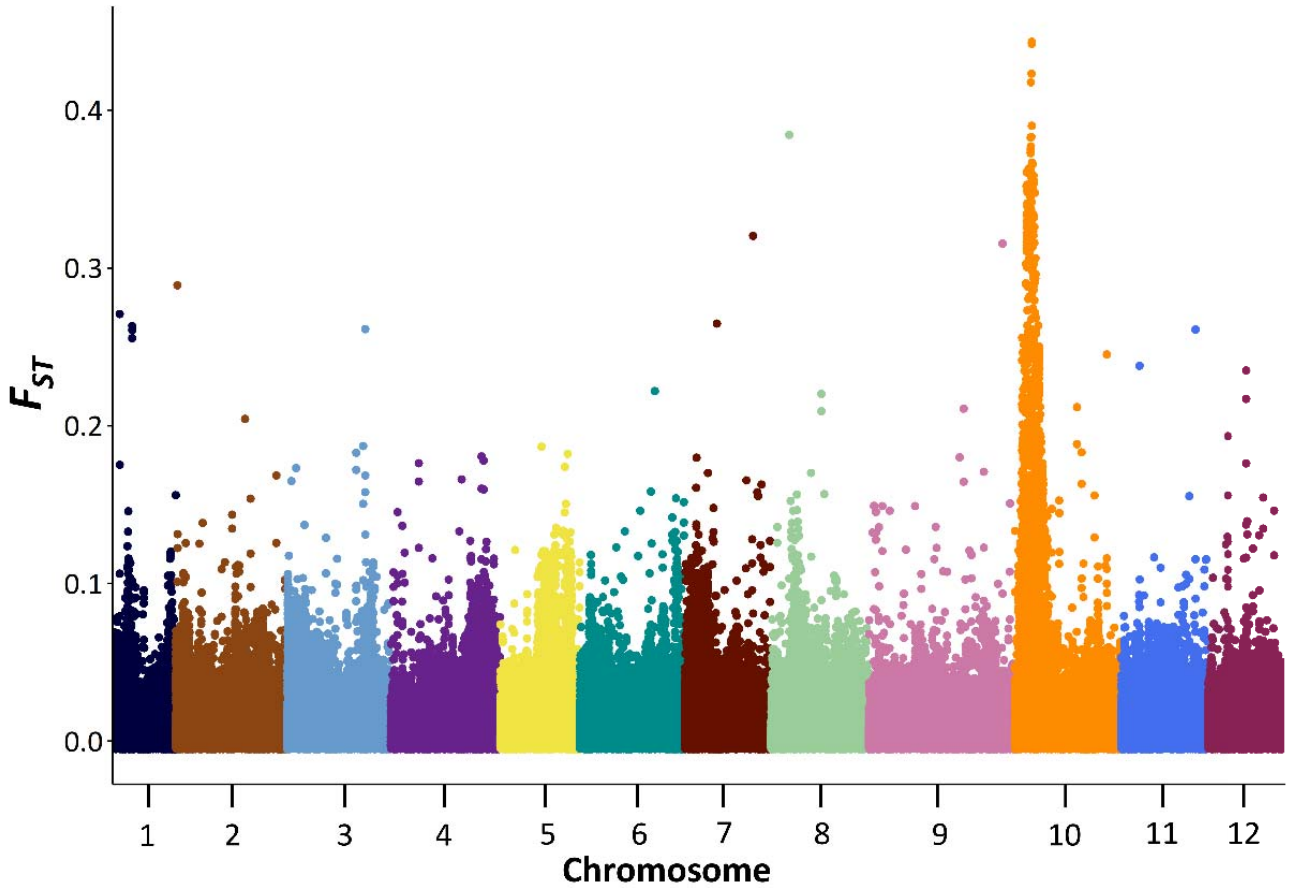


Fig. 7: Genome-wide F_{ST} values based in European beech (*Fagus sylvatica* L.) comprising 100 individuals adult beech trees per stand in five locations sampled in the South-Eastern Romanian Carpathians along an altitudinal gradient associated with precipitation and temperature.

Most GWAS for stomatal density identified several genes or QTL (Quantitative Trait Loci) (Bergmann et al. 2004; Gailing et al. 2008; McKown et al. 2019; Chhetri et al. 2020; Li et al. 2023). Li et al. (2023) also identified *Aluminum-activated malate transporter 12 (ALMT12)* in a GWAS and a co-expression network analysis in *Populus*. *ALMT12* belongs to aluminum-activated malate transporter (ALMT) family (Li et al. 2023). Based on our GWAS and the literature review, we can assume stomatal density has an oligogenic trait architecture and the identified region on chromosome 10 from ~4.89 to ~13.67 Mb seems to be very important in local adaptation. High peak SNP associations and an oligogenic trait architecture for stomatal density were also reported by McKown et al. (2019) and Chhetri et al. (2020) in *Populus trichocarpa* Hook. A linear regression of the different genotypes observed at the

SNP on chromosome 10 at ~7.82 Mb and stomatal density is available in supplemental Figure S25. The coefficient of linear regression calculated was considerably high (R^2=0.2353, p ~0).

All gene variants except for *Bhaga_10.g1674* were annotated based on other tree species of the family Fagaceae, which were *Quercus suber* L. and *Quercus lobata* Née (NCBI 2024). This makes the annotated gene variant results more reliable. Unfortunately, these genes were not studied in *Q. suber* L. and *Q. lobata* Née. *Bhaga_10.g1674* was annotated based on *Prunus dulcis* Batsch (NCBI 2024).

GWAS results for leaf nitrogen and carbon varied across traits and year of sampling

In total, 11 markers were identified as being significantly associated with leaf nitrogen content and C/N ratio measured in 2020, 2021 and 2022. Eight of these markers were < 20 kb from coding regions of annotated genes (Table 2). In total, eight gene annotations for the ten markers were available (Table 2). We only found one underlying gene that was described in the literature. The gene variant *Bhaga_2.g3457* was annotated as *abscisic-aldehyde oxidase-like* (NCBI 2024), which was observed to be related to drought stress (Little and Eidt 1968). A marker on chromosome 2 at ~ 31.8 Mb, near the abscisic aldehyde oxidase-like gene, was associated with leaf nitrogen content in 2021.

The gene *abscisic-aldehyde oxidase-like* is responsible for oxidation of abscisic aldehyde, which is the last step of abscisic acid (ABA) biosynthesis (Seo et al. 2000). ABA delays bud break in forest tree species and also inhibits simultaneously transpiration (Little and Eidt 1968). Peuke et al. (2002) observed an increase in ABA under stress in European beech. As mentioned before, ABA also plays an important role in stomatal closure (Chen et al. 2022). Lu et al. (2015) observed a connection between C/N ratio and ABA signaling pathways. Reduced leaf nitrogen supply can increase sensitivity of stomata (Gessler et al. 2017).

There is no consistent correlation trend between leaf nitrogen measured in 2021 and MAF (supplemental Fig. S11). MAF is increased in Lempes, Tampa (stands at low altitude) and Ruia (at high altitude), and decreased in Lupului and Solomon (supplemental Fig. S11). Leaf nitrogen content is decreased in Lempes, Tampa (stands at low altitude) and Ruia (at high altitude), and increased in Lupului and Solomon (supplemental Fig. S11). The phenotypic variability was observed to be much higher for traits like leaf nitrogen content and C/N ratio than for stomatal

density (De la Torre et al. 2022). More stands along the altitudinal gradient and individuals per stand may have been useful to study polygenic traits with a high phenotypic plasticity like leaf nitrogen content and C/N ratio. The R^2 value of the genotypes observed at the marker on chromosome 3 at ~31.804 Mb and leaf nitrogen content measured in 2021 was 0.0069.

Other genes involved in ABA biosynthesis were *ASPG1* (*Bhaga_10.g1462*) and *BONZAI1* (*Bhaga_10.g1651*), which are close to or overlapping with SNP markers associated with stomatal density. Given the importance of ABA biosynthesis in drought tolerance, the marker on chromosome 2 at ~31.8 Mb is a promising candidate SNP for drought stress tolerance.

No SNP marker was associated with more than one trait. However, associations for stomatal density and $\delta^{13}C$ were observed in the same region on chromosome 10. We would have expected at least a few overlapping markers between leaf nitrogen content and C/N ratio, since C/N ratio is calculated based on the leaf nitrogen content. Also, no consistent association of markers with the same trait measured in different years was found. The correlation of C/N ratio measured in different years was relatively high with a correlation coefficient of 0.497 ($p \leq 0.0001$), 0.460 ($p \leq 0.0001$) and 0.510 ($p \leq 0.0001$) (supplemental Fig. S2). The correlation of leaf nitrogen content measured in different years was similar (supplemental Fig. S2). Phenotypic variability of these traits may be too high to detect SNP-trait associations across years (De la Torre et al. 2022), which can also be observed in supplemental Figure S1. Additionally, leaf nitrogen content (N) and the C/N ratio are impacted by complex physiological mechanisms (Gessler et al. 2017). These complex physiological mechanisms depend on the C and N availability before the drought event (Gessler et al. 2017). Lower nutrient availability at the beginning of drought stress leads to a higher C/N ratio during drought, while higher nutrient availability leads to a lower C/N ratio during drought (Gessler et al. 2017). Short and intense drought stress usually causes higher nitrogen depletion, while long drought stress causes C starvation (Gessler et al. 2017). These complex physiological mechanisms may make association studies for these traits even more complicated.

Two SNP markers were associated with $\delta^{13}C$ measured in 2020 at the permutation derived significance levels of $p \leq 0.000001$. Only two SNP markers were identified as being associated with $\delta^{13}C$ measured in 2021 and 2022. In the GWAS for $\delta^{13}C$ measured in 2020 and 2022, we observed a peak on chromosome 10 that is overlapping with the

peak observed in the GWAS for stomatal density on chromosome 10 from ~4.89 to ~13.67 Mb (Fig. 1 and Fig. 2). This observation is meaningful as $\delta^{13}C$ captures changes in carbon isotope composition due to stomatal closure due to reduced soil water availability and plant water status (Fotelli et al. 2003).

$\hat{G}$ results for $\delta^{13}C$

The fact that only in one year two SNPs were associated with $\delta^{13}C$ does not mean, there is no genetic basis for this trait. Previous studies showed, that it is difficult to identify significant markers for highly polygenic traits with infinitesimal sites contributing to the genetic variation (Beissinger et al. 2018; Zeng et al. 2018; Morales et al. 2021; Mahmoud et al. 2023). Therefore, we additionally applied the polygenic test $\hat{G}$ to our data set. According to $\hat{G}$, $\delta^{13}C$ in 2020 and 2021, was under highly significant polygenic selection (p ~0) (see supplemental Table S16). $\delta^{13}C$ decreased with increasing precipitation along the altitudinal gradient (p ~0) in the two years (2020; 2021). $\hat{G}$ tested for $\delta^{13}C$ measured in 2022 was not significant (see supplemental Table S16). The $\hat{G}$ results for 2020 and 2021 are in concordance with the literature. $\delta^{13}C$, which can be used as an indicator of plant water potential, is increased (or less negative) under heat and drought stress. This leads to reduced water loss under heat stress, which was also observed in previous studies (Querejeta et al. 2003; Fotelli et al. 2003).

In the GWAS for $\delta^{13}C$ measured in 2020 and 2022, a peak on chromosome 10 which overlapped with the peak observed in the GWAS for stomatal density chromosome 10 from ~4.89 to ~13.67 Mb (Fig. 1 and Fig. 2). To summarize, we observed the clearest results for $\delta^{13}C$ measured in 2020, with signals of positive polygenic selection and two significant SNP markers near multiple candidate SNPs for stomatal density. The MAF observed at these SNP markers on chromosome 10 at ~ 10.9071 Mb and at ~10.9073 Mb correlated positively with $\delta^{13}C$ measured in 2020 at $|r|$ ~ 0.4910 (p-value = 0.08) and $|r|$ ~ 0.4843 (p-value = 0.06), but this correlation was not significant.

Conclusions

Our research gives insights into the genetic architecture of different drought stress traits. We identified several very interesting candidate SNPs for stomatal density. Most of these candidate SNPs are located in one wide region on chromosome 10 from ~4.89 to ~13.67 Mb. Within this region, we found based on a literature review, annotations for five genes with a direct impact on stomatal density, which comprise *aluminum-activated malate transporter 2-like*,

transcription factor MYB35-like, ATP-dependent Clp protease adapter protein ClpS1, Aspartic protease in guard cell 1 and *BONZAI1*. The minor allele frequencies observed at the candidate SNP markers in this region are positively correlated with stomatal density, altitude and precipitation and negatively with temperature. Therefore, we can deduce that this entire region is important for local adaptation. Furthermore, we identified one interesting candidate SNP associated with leaf nitrogen content. For $\delta^{13}C$ measured in one year, we identified two significant SNP marker associations, also located within the wide peak region on chromosome 10 from ~4.89 to ~13.67 Mb associated with stomatal density. We also observed signals of highly polygenic selection for $\delta^{13}C$. In conclusion, our research demonstrates that polygenic tests like $\hat{G}$ and GWAS can complement each other. $\hat{G}$ focuses on the small effect loci and uncovers their combined effect on a trait. Our GWAS worked very well regarding the identification of candidate SNP markers for stomatal density. We also identified a few SNP markers associated with leaf nitrogen content and C/N ratio. Unfortunately, there was no overlap between markers associated with the same trait measured in different years. Our GWAS did return only a limited number of candidate SNP markers associated with $\delta^{13}C$ in one year, but the polygenic test $\hat{G}$ returned signals of polygenic selection for this trait. Our results suggest that traits like $\delta^{13}C$ are polygenic, whereas stomatal density is influenced by a few larger-effect loci. Therefore, genomic prediction seems to be a promising tool for highly polygenic traits like $\delta^{13}C$. When traits like stomatal density are under selection, marker-assisted selection (MAS) of our few candidate SNPs for stomatal density may be sufficient. But for future genomic prediction or $\hat{G}$ studies, we would recommend the usage of larger data sets in terms of individuals, because we observed low predictive abilities in our cross validations based on this data set.

Availability of data and materials

All the data and statistics about the present study have been included in the current manuscript in the form of figures and tables or in the supplemental material. Raw data will be publicly available at figshare. Our analysis code is publicly available on github: https://github.com/milaleonie/DroughtMarkersProject.

Acknowledgements

We would like to thank all people involved in this project. We are thanking Jonah Fels, Jürgen Scheuring, Nikolas Heeger, Luca Schwendemann, Merle Kleikamp, Timo Unbehau, Jonas Daniel, Katharina Ziesing, Côme Sylvain-

dutillow, Tania Dominguez-Flores, Christoph Leuschner, Katharina B. Budde, Mehdi Ben Targem, Thi Mai Trang Schneider, Mihnea Ciocirlan and Elena Ciocirlan.

Funding

The used data comes from the “Droughtmarkers” project (Reference numbers: 2218WK43B4, 2218WK43A4). This work was financially supported by the Federal Ministry of Food and Agriculture – FNR-Waldklimafonds. The drought markers project is a collaboration between the University of Goettingen (Germany), the HAWK (Germany), the NW-FVA (Germany) and Transilvania University of Brasov (Romania). We utilized the computational resources of the University of Goettingen’s GWDG. Open Access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Alves FC, Balmant KM, Resende MFR et al. (2020) Accelerating forest tree breeding by integrating genomic selection and greenhouse phenotyping. Plant Genome 13: e20048. https://doi.org/10.1002/tpg2.20048

Beissinger T, Kruppa J, Cavero D et al. (2018) A simple test identifies selection on complex traits. Genetics 209:321–333. https://doi.org/10.1534/genetics.118.300857

Berg JJ, Coop G (2014) A population genetic signal of polygenic adaptation. PLoS Genet 10:e1004412. https://doi.org/10.1371/journal.pgen.1004412

Bergmann DC, Lukowitz W, Somerville CR (2004) Stomatal development and pattern controlled by a MAPKK kinase. Science 304:1494–1497. https://doi.org/10.1126/science.1096014

Bertolino LT, Caine RS, Gray JE (2019) Impact of stomatal density and morphology on water-use efficiency in a changing world. Front Plant Sci 10:225. https://doi.org/10.3389/fpls.2019.0022

Cappa EP, El-Kassaby YA, Garcia MN et al. (2013) Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. PLoS One 8:e81267. https://doi.org/10.1371/journal.pone.0081267

Climate data store (CDS). Copernicus Climate Change Service (C3S) European Centre for Medium-Range Weather Forecasts (ECMWF). [cited 2024 September 09]. Available from: https://cds.climate.copernicus.eu/

Chen Z, Li S, Wan X et al. (2022) Strategies of tree species to adapt to drought from leaf stomatal regulation and stem embolism resistance to root properties. Front Plant Sci 13:926535. https://doi.org/10.3389/fpls.2022.926535

Chen K, Gao J, Sun S et al. (2020) BONZAI proteins control global osmotic stress responses in plants. Curr Biol 30:4815-4825.e4. https://doi.org/10.1016/j.cub.2020.09.016

Chen Z-Q, Baison J, Pan J et al. (2018) Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. BMC Genomics 19:946. https://doi.org/10.1186/s12864-018-5256-y

Chhetri HB, Furches A, Macaya-Sanz D et al. (2020) Genome-wide association study of wood anatomical and morphological traits in *Populus trichocarpa*. Front Plant Sci 11:545748. https://doi.org/10.3389/fpls.2020.545748

Cingolani P, Platts A, Le Wang L et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92. https://doi.org/10.4161/fly.19695

Cortés AJ, Restrepo-Montoya M, Bedoya-Canas LE (2020) Modern strategies to assess and breed forest tree adaptation to changing climate. Front Plant Sci 11:583323. https://doi.org/10.3389/fpls.2020.583323

Cumbie WP, Eckert A, Wegrzyn J et al. (2011) Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. Heredity (Edinb) 107:105–114. https://doi.org/10.1038/hdy.2010.168

Dabravolski SA, Isayenkov SV (2023) Recent updates on ALMT transporters' physiology, regulation, and molecular evolution in plants. Plants (Basel) 12. https://doi.org/10.3390/plants12173167

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Degen B, Müller NA (2023) A simulation study comparing advanced marker-assisted selection with genomic selection in tree breeding programs. G3 (Bethesda) 13. https://doi.org/10.1093/g3journal/jkad164

De La Torre AR, Sekhwal MK, Puiu D et al. (2022) Genome-wide association identifies candidate genes for drought tolerance in coast redwood and giant sequoia. Plant J 109:7–22. https://doi.org/10.1111/tpj.15592

Deli J, Matus Z, Tóth G (2000) Carotenoid composition in the fruits of *Asparagus officinalis*. J Agric Food Chem 48:2793–2796. https://doi.org/10.1021/jf991243h

Deng YJ, Duan AQ, Liu H et al. (2024) Generation of high purity capsanthin and capsorubin through synthetic metabolic engineering in carrot germplasm. J Exp Bot 75:7202–7216. https://doi.org/10.1093/jxb/erae351

Diao S, Ding X, Luan Q et al. (2024). Development of 51K liquid phased probe array for loblolly and slash pines and its application to GWAS of slash pine breeding population. Industrial Crops and Products: 216(118777). https://doi.org/10.1016/j.indcrop.2024.118777.

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R Package rrBLUP. Plant Genome 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Feng Y, Cohen SN (2000) Unpaired terminal nucleotides and 5′ monophosphorylation govern 3′ polyadenylation by *Escherichia coli* poly(A) polymerase I. Proc Natl Acad Sci USA: 6415–6420

Fisher RA (1918) XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. Earth and Environmental Science Transactions of The Royal Society of Edinburgh 52:399–433

Fotelli MN, Rennenberg H, Holst T et al. (2003) Carbon isotope composition of various tissues of beech (*Fagus sylvatica*) regeneration is indicative of recent environmental conditions within the forest understorey. New Phytol 159:229–244. https://doi.org/10.1046/j.1469-8137.2003.00782.x

Frangedakis E, Yelina NE, Billakurthi K et al. (2024) MYB-related transcription factors control chloroplast biogenesis. Cell 187:4859-4876.e22. https://doi.org/10.1016/j.cell.2024.06.039

Franks PJ, Drake PL, Beerling DJ (2009) Plasticity in maximum stomatal conductance constrained by negative correlation between stomatal size and density: an analysis using *Eucalyptus globulus*. Plant Cell & Environment 32:1737–1748. https://doi.org/10.1111/j.1365-3040.2009.002031.x

Gailing O, Langenfeld-Heyser R, Polle A, Finkeldey R (2008) Quantitative trait loci affecting stomatal density and growth in a *Quercus robur* progeny: implications for the adaptation to changing environments. Global Change Biology 14:1934–1946. https://doi.org/10.1111/j.1365-2486.2008.01621.x

Gangadhar BH, Mishra RK, Kappachery S, Baskar V, Venkatesh J, Nookaraju A, Thiruvengadam M (2021) Enhanced thermo-tolerance in transgenic potato (*Solanum tuberosum* L.) overexpressing hydrogen peroxide-producing germin-like protein (GLP). Genomics 113(5): 3224-3234

Gao F, Dubos C (2024) The Arabidopsis bHLH transcription factor family. Trends Plant Sci 29:668–680. https://doi.org/10.1016/j.tplants.2023.11.022

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv preprint; arXiv:1207.3907 [q-bio.GN]. https://doi.org/10.48550/arXiv.1207.3907

Gou M, Zhang Z, Zhang N et al. (2015) Opposing effects on two phases of defense responses from concerted actions of HEAT SHOCK COGNATE70 and BONZAI1 in *Arabidopsis*. Plant Physiol 169:2304–2323. https://doi.org/10.1104/pp.15.00970

Grigoriadou-Zormpa O, Wilhelmi S, Vucetic B, Ciocîrlan MIC, Mueller M, Ciocîrlan E, Lucian Curtu A, Targem MB, H Wildhagen H, Gailing O, Budde KB. Differences in fine-scale spatial genetic structure of European beech populations along elevational gradients. 2024; PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-4559673/v1].

Hamanishi ET, Thomas BR, Campbell MM (2012) Drought induces alterations in the stomatal development program in *Populus*. J Exp Bot 63:4959–4971. https://doi.org/10.1093/jxb/ers177

Hancock AM, Witonsky DB, Gordon AS et al. (2008) Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet 4:e32. https://doi.org/10.1371/journal.pgen.0040032

IGA Technology Services Srl, Udine, Italy. https://igatechnology.com/.

Josephs EB, Berg JJ, Ross-Ibarra J et al. (2019) Detecting adaptive differentiation in structured populations with genomic data and common gardens. Genetics 211:989–1004. https://doi.org/10.1534/genetics.118.301786

Lauer E, Holland J, Isik F (2022) Prediction ability of genome-wide markers in *Pinus taeda* L. within and between population is affected by relatedness to the training population and trait genetic architecture. G3 (Bethesda) 12. https://doi.org/10.1093/g3journal/jkab405

Lazic D, Geßner C, Liepe KJ et al. (2024) Genomic variation of European beech reveals signals of local adaptation despite high levels of phenotypic plasticity. Nat Commun 15:8553. https://doi.org/10.1038/s41467-024-52933-y

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li L, Jin Z, Huang R et al. (2023) Leaf physiology variations are modulated by natural variations that underlie stomatal morphology in *Populus*. Plant Cell Environ 46:150–170. https://doi.org/10.1111/pce.14471

Lindman S, Hernandez-Garcia A, Szczepankiewicz O et al. (2010) In vivo protein stabilization based on fragment complementation and a split GFP system. Proc Natl Acad Sci USA 107:19826–19831. https://doi.org/10.1073/pnas.1005689107

Little CHA, Eidt DC (1968) Effect of abscisic acid on budbreak and transpiration in woody species. Nature 220: 498–499

Lu Y, Sasaki Y, Li X et al. (2015) ABI1 regulates carbon/nitrogen-nutrient signal transduction independent of ABA biosynthesis and canonical ABA signalling pathways in *Arabidopsis*. J Exp Bot 66:2763–2771. https://doi.org/10.1093/jxb/erv086

Mahmoud M, Tost M, Ha NT et al. (2023) Ghat: an R package for identifying adaptive polygenic traits. G3 (Bethesda) 13. https://doi.org/10.1093/g3journal/jkac319

Manosalva PM, Davidson RM, Liu B et al. (2009) A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. Plant Physiol 149:286–296. https://doi.org/10.1104/pp.108.128348

McKown AD, Klápště J, Guy RD et al. (2019) A role for SPEECHLESS in the integration of leaf stomatal patterning with the growth vs disease trade-off in poplar. New Phytol 223:1888–1903. https://doi.org/10.1111/nph.15911

Meger J, Ulaszewski B, Pałucka M et al. (2024) Genomic prediction of resistance to *Hymenoscyphus fraxineus* in common ash (*Fraxinus excelsior* L.) populations. Evol Appl 17:e13694. https://doi.org/10.1111/eva.13694

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Mishra B, Ulaszewski B, Meger J et al. (2021) A chromosome-level genome assembly of the European beech (*Fagus sylvatica*) reveals anomalies for organelle DNA integration, repeat content and distribution of SNPs. Front Genet 12:691058. https://doi.org/10.3389/fgene.2021.691058

Morales L, Michel S, Ametz C et al. (2021) Genomic signatures of selection for resistance to stripe rust in Austrian winter wheat. Theor Appl Genet 134:3111–3121. https://doi.org/10.1007/s00122-021-03882-3

Müller M, Leuschner C, Weithmann G et al. (2024) A genome-wide genetic association study reveals SNPs significantly associated with environmental variables and specific leaf area in European beech. Physiol Plant 176:e14334. https://doi.org/10.1111/ppl.14334

National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information. [cited 2024 December 09]. Available from: https://www.ncbi.nlm.nih.gov/.

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111–122. https://doi.org/10.1038/nrg2931

Negi J, Hashimoto-Sugimoto M, Kusumi K et al. (2014) New approaches to the biology of stomatal guard cells. Plant Cell Physiol 55:241–250. https://doi.org/10.1093/pcp/pct145

Nishimura K, Apitz J, Friso G, Kim J, Ponnala L, Grimm B, van Wijk KJ (2015) Discovery of a unique Clp component, ClpF, in chloroplasts: A proposed binary ClpF-ClpS1 adaptor complex functions in substrate recognition and delivery. The Plant Cell 27(10): 2677–2691

Nishihama R, Soyano T, Ishikawa M, Araki S, Tanaka H, Asada T, Irie K, Ito M, Terada M, H Banno, Yamazaki Y, Machida Y (2002) Expansion of the cell plate in plant cytokinesis requires a kinesin-like protein/MAPKKK complex. Cell 109(1):87-99

Nystedt B, Street NR, Wetterbom A et al. (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584. https://doi.org/10.1038/nature12211

Peuke AD, Schraml C, Hartung W et al. (2002) Identification of drought-sensitive beech ecotypes by physiological parameters. New Phytol 154:373–387. https://doi.org/10.1046/j.1469-8137.2002.00400.x

Pfenninger M, Reuss F, Kiebler A et al. (2021) Genomic basis for drought resistance in European beech forests threatened by climate change. Elife 10. https://doi.org/10.7554/eLife.65532

Piano D, Cocco E, Guadalupi G et al. (2019) Characterization under quasi-native conditions of the capsanthin/capsorubin synthase from *Capsicum annuum* L. Plant Physiol Biochem 143:165–175. https://doi.org/10.1016/j.plaphy.2019.09.007

Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795

Querejeta JI, Barea JM, Allen MF et al. (2003) Differential response of delta13C and water use efficiency to arbuscular mycorrhizal infection in two aridland woody plant species. Oecologia 135:510–515. https://doi.org/10.1007/s00442-003-1209-4

R Core Team (2024) R: A Language and Environment for Statistical. https://CRAN.R-project.org/.

R Core Team (2025) R: A Language and Environment for Statistical. https://CRAN.R-project.org/.

Racimo F, Berg JJ, Pickrell JK (2018) Detecting polygenic adaptation in admixture graphs. Genetics 208:1565–1584. https://doi.org/10.1534/genetics.117.300489

Rellstab C, Gugerli F, Eckert AJ et al. (2015) A practical guide to environmental association analysis in landscape genomics. Mol Ecol 24:4348–4370. https://doi.org/10.1111/mec.13322

Ruehr NK, Grote R, Mayr S et al. (2019) Beyond the extreme: recovery of carbon and water relations in woody plants following heat and drought stress. Tree Physiol 39:1285–1299. https://doi.org/10.1093/treephys/tpz032

Sasabe M, Boudolf V, Veylder L de et al. (2011) Phosphorylation of a mitotic kinesin-like protein and a MAPKKK by cyclin-dependent kinases (CDKs) is involved in the transition to cytokinesis in plants. Proc Natl Acad Sci USA 108:17844–17849. https://doi.org/10.1073/pnas.1110174108

Scaglione D, Pinosio S, Marroni F et al. (2019) Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. Ann Bot 124:543–552. https://doi.org/10.1093/aob/mcz054

Seo M, Koiwai H, Akaba S et al. (2000) Abscisic aldehyde oxidase in leaves of *Arabidopsis thaliana*. Plant J 23:481–488. https://doi.org/10.1046/j.1365-313x.2000.00812.x

Shen W, Yao X, Ye T et al. (2018) *Arabidopsis* aspartic protease ASPG1 affects seed dormancy, seed longevity and seed germination. Plant Cell Physiol 59:1415–1431. https://doi.org/10.1093/pcp/pcy070

Shin J-H, Blay S, Graham J et al. (2006) LDheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. J Stat Soft 16. https://doi.org/10.18637/jss.v016.c03

Valadon L, Mummery RS (1977) Carotenoids of lilies and of red pepper: Biogenesis of Capsanthin and Capsorubin. Zeitschrift für Pflanzenphysiologie 82:407–416. https://doi.org/10.1016/S0044-328X(77)80004-4

Vidya E, Duchaine TF (2022) Eukaryotic mRNA decapping activation. Front Genet 13:832547. https://doi.org/10.3389/fgene.2022.832547

Visscher PM, Wray NR, Zhang Q et al. (2017) 10 years of GWAS discovery: Biology, function, and translation. Am J Hum Genet 101:5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Voith von Voithenberg L, Park J, Stübe R et al. (2019) A novel prokaryote-type ECF/ABC transporter module in chloroplast metal homeostasis. Front Plant Sci 10:1264. https://doi.org/10.3389/fpls.2019.01264

Wang B, Yu J, Luo M et al. (2024) Aspartic proteases gene family: Identification and expression profiles during stem vascular development in tobacco. Int J Biol Macromol 279:135016. https://doi.org/10.1016/j.ijbiomac.2024.135016

Werner RA, Bruch BA, Brand WA (1999) ConFlo III - an interface for high precision δ13C and δ15N analysis with an extended dynamic range. Rapid Commun Mass Spectrom 13:1237–1241. https://doi.org/10.1002/(SICI)1097-0231(19990715)13:13<1237:AID-RCM633>3.0.CO;2-C

Wright S (1931) Evolution in Mendelian populations. Genetics 16(2): 97–159.

Yao X, Xiong W, Ye T et al. (2012) Overexpression of the aspartic protease ASPG1 gene confers drought avoidance in *Arabidopsis*. J Exp Bot 63:2579–2593. https://doi.org/10.1093/jxb/err433

Yücedağ C, Sanders J, Musah M et al. (2019) Stomatal density in *Quercus petraea* and *Q. robur* natural populations in Northern Turkey. Dendrobiology 81:58–64. https://doi.org/10.12657/denbio.081.007

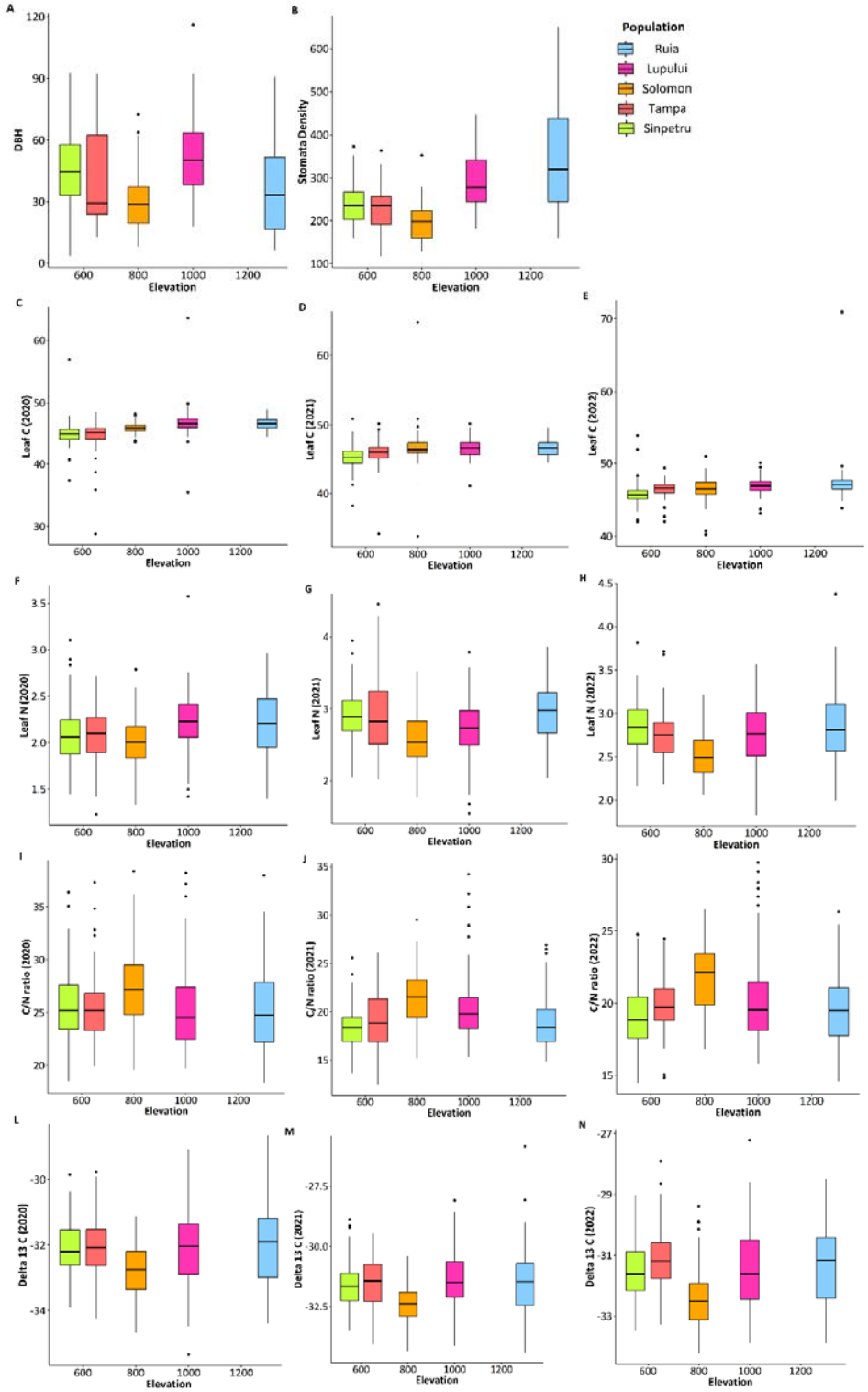
Zeng J, de Vlaming R, Wu Y et al. (2018) Signatures of negative selection in the genetic architecture of human complex traits. Nat Genet 50:746–753. https://doi.org/10.1038/s41588-018-0101-4

Zhang M, Wei F, Guo K et al. (2016) A novel FC116/BC10 mutation distinctively causes alteration in the expression of the genes for cell wall polymer synthesis in rice. Front Plant Sci 7:1366. https://doi.org/10.3389/fpls.2016.01366

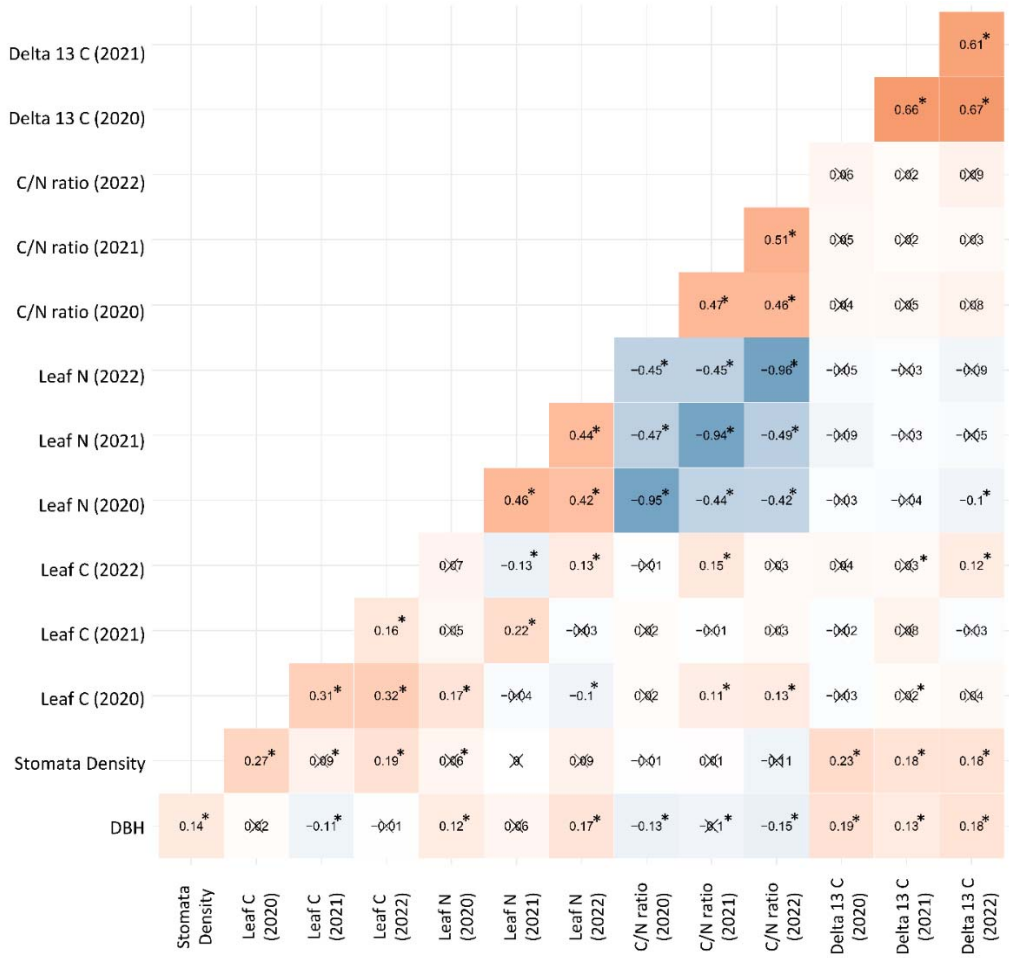
Zhang Y, Yin B, Zhang J et al. (2019) Histone deacetylase HDT1 is involved in stem vascular development in *Arabidopsis*. Int J Mol Sci 20. https://doi.org/10.3390/ijms20143452

Supplementary files

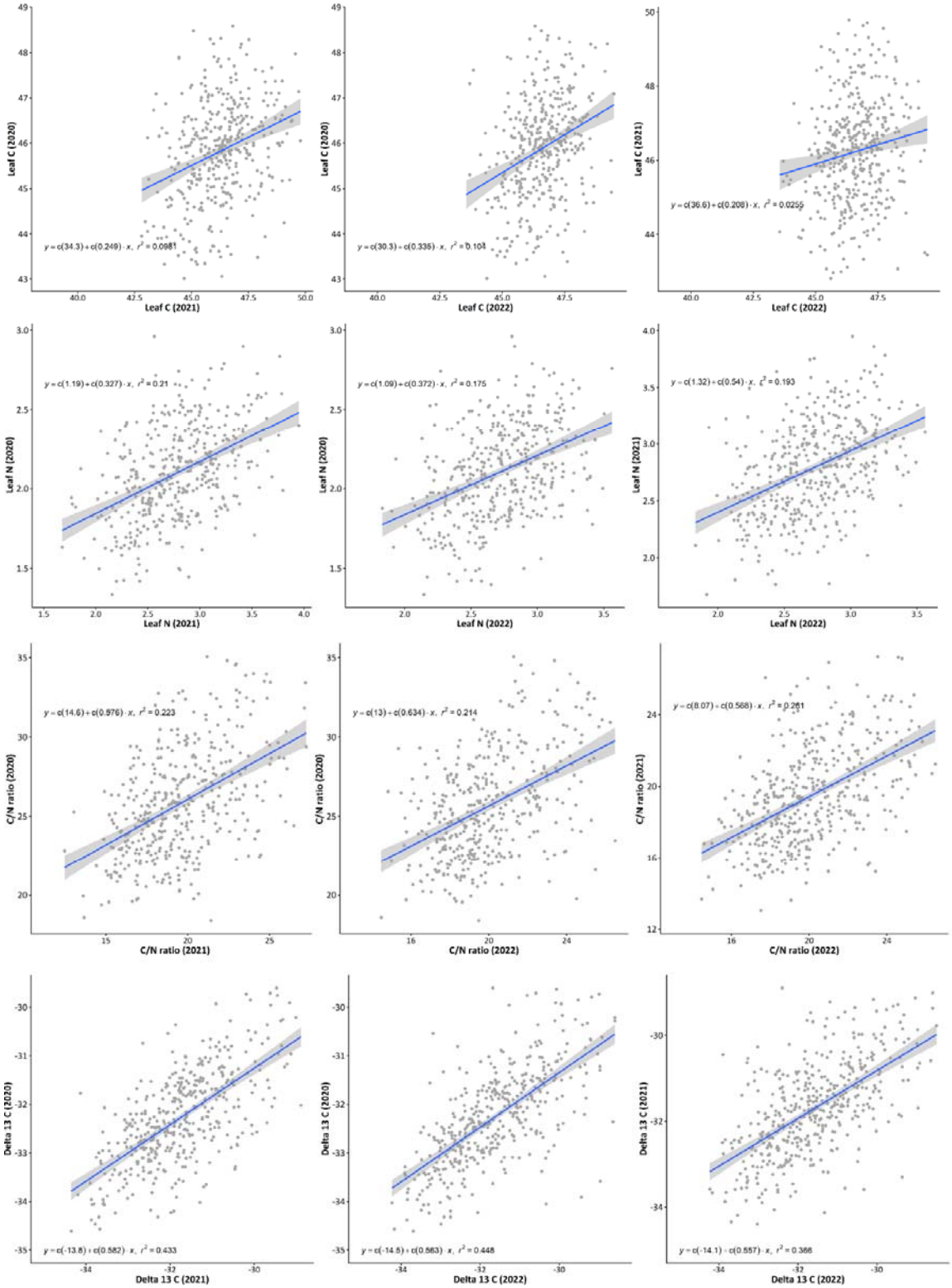
S1: Phenotypic trait measurements: diameter at breast height (DBH), stomatal density, leaf carbon content measured in 2020, 2021 2022, leaf nitrogen content measured in 2020, 2021 2022, C/N ratio calculated for 2020, 2021 2022 and water use efficiency as $\delta^{13}C$ measured in 2020, 2021 2022.



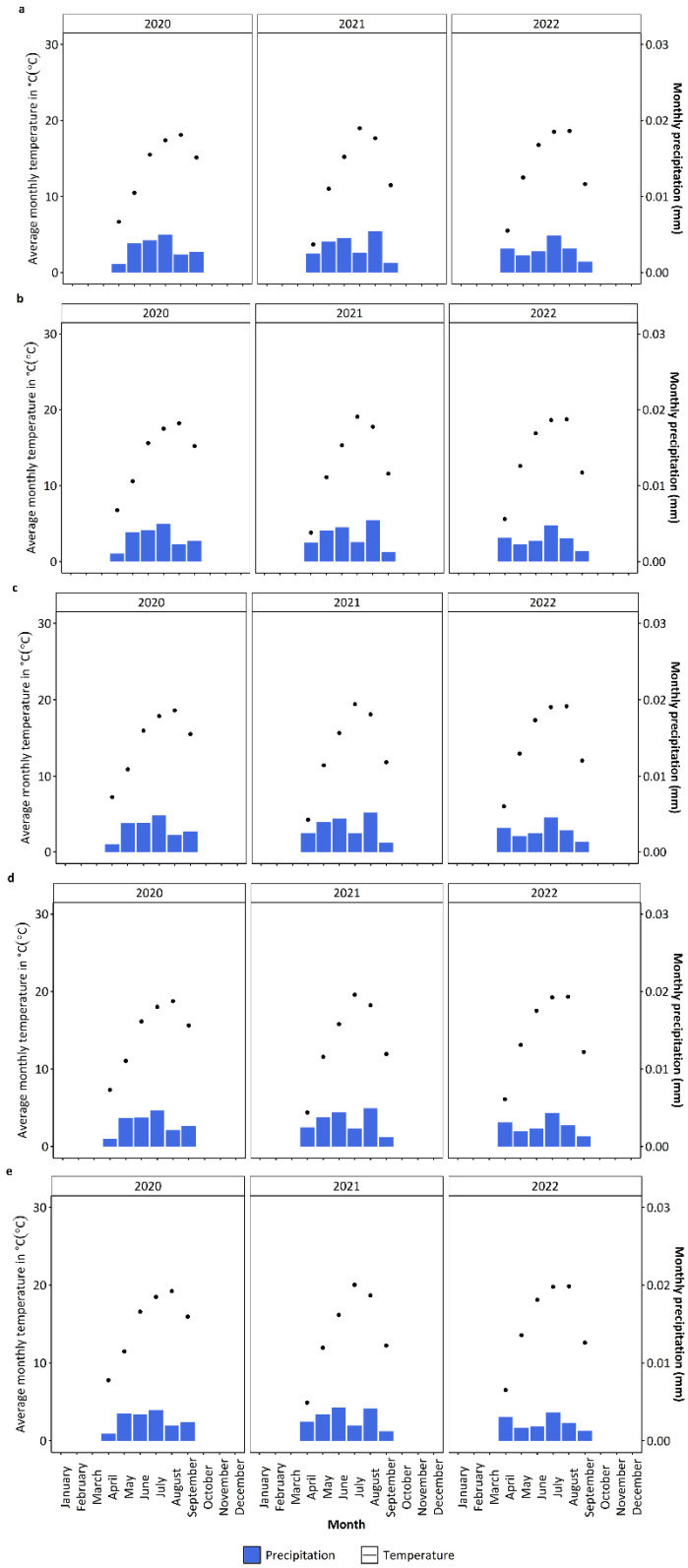
S2: Correlations of all phenotypic traits and their correlation coefficients ($p \leq 0.05$).



S3: Correlations of the traits measured in different years for leaf carbon content measured in 2020, 2021 2022, leaf nitrogen content measured in 2020, 2021 2022, C/N ratio calculated for 2020, 2021 2022 and water use efficiency as $\delta^{13}C$ measured in 2020, 2021 2022.



S4: Average monthly temperature in °C and monthly precipitation (mm) measured during the sampling season from April to September in 2020, 2021 and 2022 in Ruia (a), Lupului (b), Solomon (c), Tampa (d) and Lempes (e), which were used in the environmental principal component analysis (PCA).

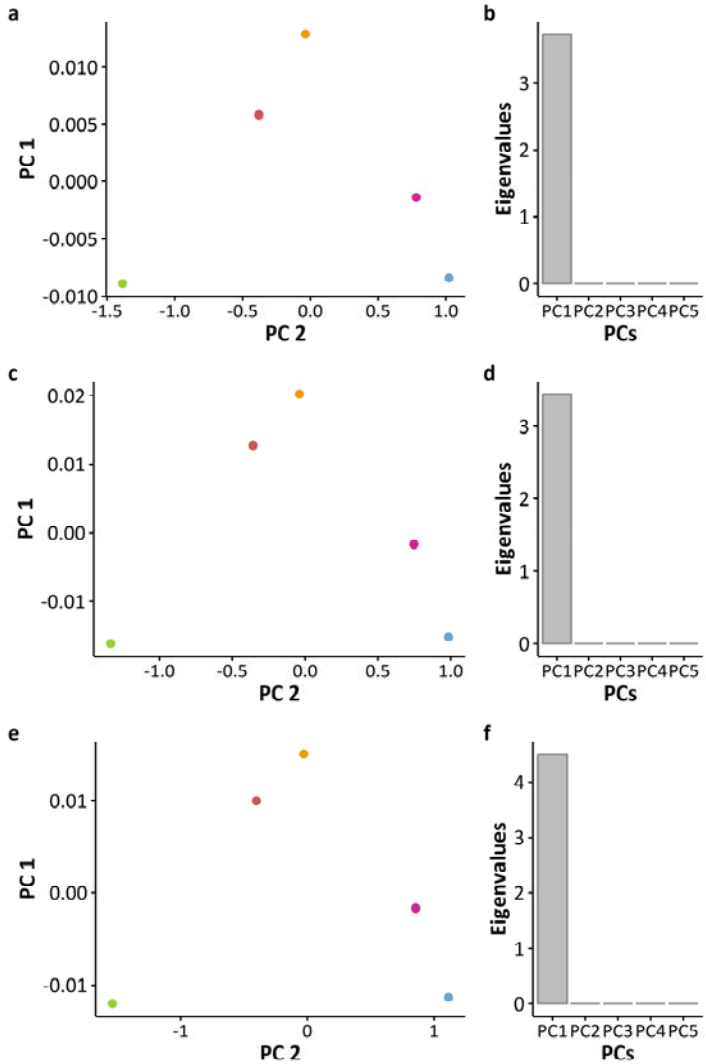


S5: Average monthly temperature in °C and monthly precipitation (mm) measured during the sampling season from April to September in 2020, 2021 and 2022 in Ruia, Lupului, Solomon, Tampa and Lempes, which were used in the environmental principal component analysis (PCA).

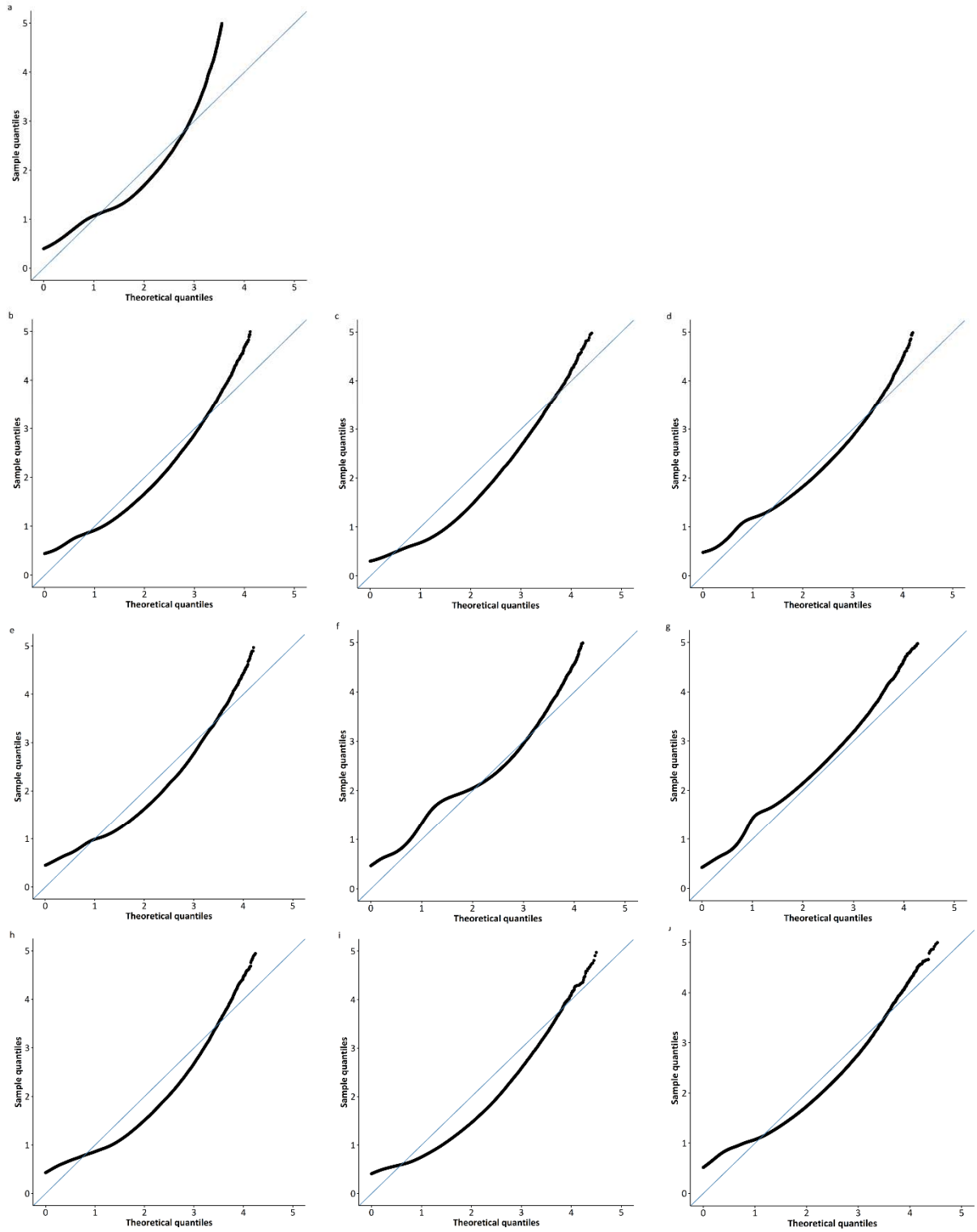
Stand	Year	Stand	Precipitation in mm	Temperature in °C
April	2020	Ruia	0.001097	6.68
April	2020	Lupului	0.001082	6.79
April	2020	Solomon	0.001027	7.16
April	2020	Tampa	0.001003	7.32
April	2020	Lempes	0.000903	7.79
May	2020	Ruia	0.003846	10.49
May	2020	Lupului	0.003833	10.58
May	2020	Solomon	0.003749	10.92
May	2020	Tampa	0.00369	11.06
May	2020	Lempes	0.003501	11.48
June	2020	Ruia	0.004246	15.52
June	2020	Lupului	0.004148	15.62
June	2020	Solomon	0.003832	15.98
June	2020	Tampa	0.00372	16.14
June	2020	Lempes	0.00336	16.60
July	2020	Ruia	0.004998	17.40
July	2020	Lupului	0.004955	17.51
July	2020	Solomon	0.004757	17.88
July	2020	Tampa	0.004615	18.03
July	2020	Lempes	0.003897	18.48
August	2020	Ruia	0.002331	18.11
August	2020	Lupului	0.002303	18.23
August	2020	Solomon	0.002197	18.62
August	2020	Tampa	0.002146	18.77
August	2020	Lempes	0.001981	19.23
April	2021	Ruia	0.002523	3.70
April	2021	Lupului	0.002513	3.82
April	2021	Solomon	0.002485	4.22
April	2021	Tampa	0.002478	4.39
April	2021	Lempes	0.002451	4.91
May	2021	Ruia	0.004061	11.02
May	2021	Lupului	0.004053	11.11
May	2021	Solomon	0.003926	11.44
May	2021	Tampa	0.003799	11.57
May	2021	Lempes	0.003399	11.95
June	2021	Ruia	0.00456	15.22
June	2021	Lupului	0.0045	15.33
June	2021	Solomon	0.004386	15.67
June	2021	Tampa	0.004396	15.79
June	2021	Lempes	0.004276	16.18
July	2021	Ruia	0.002571	18.99
July	2021	Lupului	0.002556	19.09
July	2021	Solomon	0.002439	19.44
July	2021	Tampa	0.002333	19.59
July	2021	Lempes	0.001943	20.05
August	2021	Ruia	0.005456	17.67
August	2021	Lupului	0.005416	17.77
August	2021	Solomon	0.005146	18.10
August	2021	Tampa	0.004929	18.24
August	2021	Lempes	0.004124	18.68
April	2022	Ruia	0.003118	5.52
April	2022	Lupului	0.003131	5.61
April	2022	Solomon	0.003158	5.95
April	2022	Tampa	0.003144	6.10
April	2022	Lempes	0.003019	6.53
May	2022	Ruia	0.002278	12.50

May	2022	Lupului	0.002242	12.61
May	2022	Solomon	0.002058	12.98
May	2022	Tampa	0.001948	13.12
May	2022	Lempes	0.001672	13.57
June	2022	Ruia	0.002778	16.78
June	2022	Lupului	0.002709	16.91
June	2022	Solomon	0.002456	17.34
June	2022	Tampa	0.002332	17.53
June	2022	Lempes	0.00181	18.12
July	2022	Ruia	0.004829	18.52
July	2022	Lupului	0.004786	18.64
July	2022	Solomon	0.004518	19.06
July	2022	Tampa	0.004322	19.24
July	2022	Lempes	0.003657	19.78
August	2022	Ruia	0.003133	18.63
August	2022	Lupului	0.003066	18.76
August	2022	Solomon	0.002836	19.16
August	2022	Tampa	0.002729	19.33
August	2022	Lempes	0.002308	19.85

S6: Results from the principal coordinate analysis (PCoA) on precipitation and temperature, of the sampling years 2020 (**a**) and the eigenvalues of PCs (**b**), 2021 (**c** and **d**) and 2022 (**e** and **f**). Environmental PC 1, which explained ~100% was used in the genome-wide association study (GWAS) as covariate.



S7: QQ-Plots of all phenotypic variables. The QQ plots are following this order: Stomatal density (**a**), $\delta^{13}C$ measured in 2020 (**b**), $\delta^{13}C$ measured in 2021 (**c**), $\delta^{13}C$ measured in 2022 (d), leaf nitrogen measured in 2020 (**e**), leaf nitrogen measured in 2021 (**f**), leaf nitrogen measured in 2022 (**g**), C/N ratio in 2020 (**h**), C/N ratio in 2021 (**i**), C/N ratio in 2022 (**j**).



S8: List of remaining candidate SNPs associated with stomatal density (SD) and $\delta^{13}C$ measured in 2020 with premutation derived significance levels of $p \leq 0.000001$ and their annotated gene variants, the underlying genes with their locations, functions and predicted impact (LOW or MODIFIER (not predicted)) based on SnpEff (Cingolani et al. 2012).

Trait	Chr	Position of SNP	Distance to gene variant in kb	Gene variant	Gene
SD	4	40808183	NA	NA	NA
		40808384	NA	NA	NA
		40816400	NA	NA	NA
		40818194	NA	NA	NA
		40854731	NA	NA	NA
	5	40871447	NA	NA	NA
	6	15825445	NA	NA	NA
		15825457	NA	NA	NA
		15825492	NA	NA	15825492
		15825793	NA	NA	15825793
		15825811	NA	NA	15825811
	8	7588499	2.522	Bhaga_8.g971	NA
			2.505	Bhaga_8.g971	NA
			2.206	Bhaga_8.g970	NA
	9	14893568	~-6.168 to -6.31	Bhaga_9.g1841	TYK12172.1
			~6.161	Bhaga_9.g1841	TYK12172.1
			~6.207	Bhaga_9.g1841	TYK12172.1
		5470445	NA	NA	NA
		6025501	NA	NA	NA
		6851324	NA	NA	NA
		6858935	NA	NA	NA
		6876799	NA	NA	NA
		6970293	NA	NA	NA
		7116359	~-6.0	Bhaga_10.g853	KAB1223135.1
		7116359	~-7.8	Bhaga_10.g853	KAB1223135.1
		7116359	~-14.2	Bhaga_10.g857	XP_018814658.1
		7211779	~-9.1 to 14	Bhaga_10.g871	XP_023888712.1
		7513203	NA	NA	NA
		7776590	NA	NA	NA
		7787247	NA	NA	NA
		7825443	NA	NA	NA
		8260485	NA	NA	NA
		8260499	NA	NA	NA
		8323901	NA	NA	NA
		8399501	NA	NA	NA
		8586058	NA	NA	NA
		9215197	NA	NA	NA
		9520963	NA	NA	NA
		8586058	NA	NA	NA
		9215197	NA	NA	NA
		9520963	NA	NA	NA
		10058669	NA	NA	NA
		10159262	NA	NA	NA
		10178899	NA	NA	NA
		10527710	NA	NA	NA
		10540324	NA	NA	NA
		10644942	NA	NA	NA
		10644983	NA	NA	NA
		10653218	NA	NA	NA

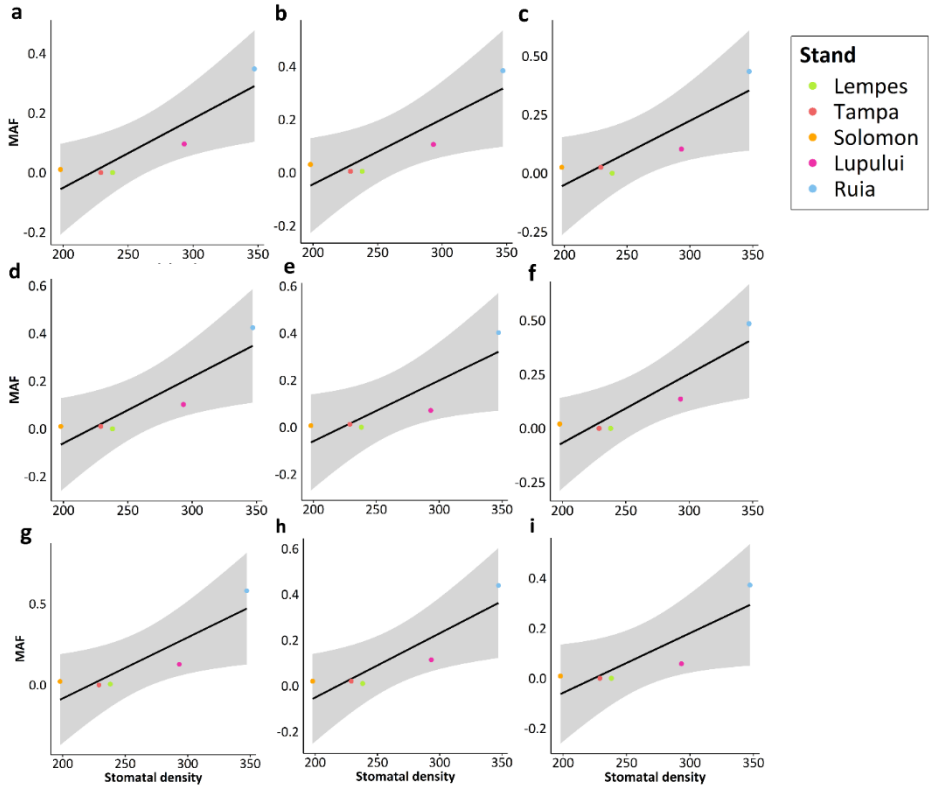
	10688216	NA	NA	NA
	10806739	NA	NA	NA
	10822774	NA	NA	NA
$\delta^{13}C$ (2020)	10907107	-3.796	Bhaga_10.g1348	XP_030940353.1
	10907267	-3.643	Bhaga_10.g1348	
SD	10909086	-1.853	Bhaga_10.g1353	
	10909126	~-18.6 to -18.4	Bhaga_10.g1353	PQQ02746.1
	10909168			
	10909306			
	10909343			
	10915528	2.844	Bhaga_10.g1350	XP_030941044.1
	11233527	~-9.6 to -11.4	Bhaga_10.g1390	XP_023922841.1
	11233527			
	11233581			
	11401610	NA	NA	NA
	11467008	NA	NA	NA
	11467225	NA	NA	NA
	11511420	~6.9 to 10.4	Bhaga_10.g1422	XP_023884902.1
		~-11.4	Bhaga_10.g1424	KAB1212694.1
	11599598	NA	NA	NA
	11634927	NA	NA	NA
	11988800	NA	NA	NA
	12053423	NA	NA	NA
	12684288	NA	NA	NA

S9: Minor allele frequency (MAF) observed in the different stands at the significant markers ($p \leq 0.000001$) on chromosome 10 associated with stomatal density and the number of heterozygous individuals (Nr of Het), number of homozygous individuals (Nr of Hom) and the total number of genotyped individuals (Nr of total ind.).

Marker	**Stand**	**MAF**	**Nr of Het.**	**Nr of Hom.**	**Nr of total ind.**	**Marker**	**MAF**	**Nr of Het**	**Nr of Hom**	**Nr of total ind.**
chr10_1090912	Sinpetru	0	0	0	95	chr10_1367781	0	0	0	67
chr10_1090912	Tampa	0	0	0	85	chr10_1367781	0	0	0	56
chr10_1090912	Solomon	0.0217	4	0	92	chr10_1367781	0.0085	1	0	59
chr10_1090912	Lupului	0.1312	19	1	80	chr10_1367781	0.0625	6	0	48
chr10_1090912	Ruia	0.4899	65	16	99	chr10_1367781	0.3661	13	14	56
chr10_1090916	Sinpetru	0	0	0	95	chr10_489837	0.02	4	0	100
chr10_1090916	Tampa	0	0	0	91	chr10_489837	0.0051	1	0	99
chr10_1090916	Solomon	0.0204	4	0	98	chr10_489837	0.0102	2	0	98
chr10_1090916	Lupului	0.1359	21	2	92	chr10_489837	0.1061	19	1	99
chr10_1090916	Ruia	0.4848	64	16	99	chr10_489837	0.2879	45	6	99
chr10_1090930	Sinpetru	0	0	0	100	chr10_631415	0	0	0	100
chr10_1090930	Tampa	0	0	0	99	chr10_631415	0	0	0	99
chr10_1090930	Solomon	0.0204	4	0	98	chr10_631415	0.0102	2	0	98
chr10_1090930	Lupului	0.1224	22	1	98	chr10_631415	0.096	19	0	99
chr10_1090930	Ruia	0.4798	65	15	99	chr10_631415	0.3485	53	8	99
chr10_1090934	Sinpetru	0	0	0	100	chr10_711635	0.005	1	0	100
chr10_1090934	Tampa	0	0	0	99	chr10_711635	0.0051	1	0	99
chr10_1090934	Solomon	0.0204	4	0	98	chr10_711635	0.0306	4	1	98
chr10_1090934	Lupului	0.1224	22	1	98	chr10_711635	0.1061	17	2	99
chr10_1090934	Ruia	0.4798	65	15	99	chr10_711635	0.3838	50	13	99
chr10_1123352	Sinpetru	0.0054	1	0	93	chr10_719880	0.015	3	0	100
chr10_1123352	Tampa	0	0	0	97	chr10_719880	0.0354	7	0	99
chr10_1123352	Solomon	0.0215	2	1	93	chr10_719880	0.0357	7	0	98
chr10_1123352	Lupului	0.1276	21	2	98	chr10_719880	0.1364	27	0	99
chr10_1123352	Ruia	0.5808	45	35	99	chr10_719880	0.3838	52	12	99
chr10_1123358	Sinpetru	0.0053	1	0	94	chr10_721177	0.0052	1	0	96

chr10_1123358	Tampa	0	0	0	97	chr10_721177	0.0152	3	0	99
chr10_1123358	Solomon	0.0215	2	1	93	chr10_721177	0.0204	4	0	98
chr10_1123358	Lupului	0.1276	21	2	98	chr10_721177	0.1042	20	0	96
chr10_1123358	Ruia	0.5808	45	35	99	chr10_721177	0.3687	51	11	99
chr10_1151142	Sinpetru	0.0643	1	4	70	chr10_755543	0	0	0	99
chr10_1151142	Tampa	0.037	0	2	54	chr10_755543	0.0253	5	0	99
chr10_1151142	Solomon	0.058	2	3	69	chr10_755543	0.0255	5	0	98
chr10_1151142	Lupului	0.2041	4	8	49	chr10_755543	0.1031	20	0	97
chr10_1151142	Ruia	0.7011	12	55	87	chr10_755543	0.4343	54	16	99
chr10_1190395	Sinpetru	0.0165	3	0	91	chr10_778592	0	0	0	100
chr10_1190395	Tampa	0.0054	1	0	93	chr10_778592	0.0101	2	0	99
chr10_1190395	Solomon	0.0054	1	0	92	chr10_778592	0.0102	2	0	98
chr10_1190395	Lupului	0.1292	21	1	89	chr10_778592	0.102	20	0	98
chr10_1190395	Ruia	0.5312	56	23	96	chr10_778592	0.4242	56	14	99
chr10_1254710	Sinpetru	0.0104	2	0	96	chr10_778597	0	0	0	100
chr10_1254710	Tampa	0.0206	4	0	97	chr10_778597	0.0102	2	0	98
chr10_1254710	Solomon	0.0206	4	0	97	chr10_778597	0.0102	2	0	98
chr10_1254710	Lupului	0.1141	17	2	92	chr10_778597	0.1031	20	0	97
chr10_1254710	Ruia	0.4394	57	15	99	chr10_778597	0.4242	56	14	99
chr10_1266723	Sinpetru	0.064	9	1	86	chr10_778746	0	0	0	96
chr10_1266723	Tampa	0.0959	10	2	73	chr10_778746	0.0102	2	0	98
chr10_1266723	Solomon	0.0769	12	0	78	chr10_778746	0.0103	2	0	97
chr10_1266723	Lupului	0.189	27	2	82	chr10_778746	0.1064	20	0	94
chr10_1266723	Ruia	0.3646	66	2	96	chr10_778746	0.4242	56	14	99
chr10_1343528	Sinpetru	0	0	0	72	chr10_782027	0	0	0	84
chr10_1343528	Tampa	0.0204	2	0	49	chr10_782027	0.0128	1	0	39
chr10_1343528	Solomon	0.0211	3	0	71	chr10_782027	0.0068	1	0	74
chr10_1343528	Lupului	0.1071	7	1	42	chr10_782027	0.0714	6	0	42
chr10_1343528	Ruia	0.2722	31	6	79	chr10_782027	0.4026	42	10	77
chr10_1367775	Sinpetru	0	0	0	64	chr10_782038	0.0101	2	0	99
chr10_1367775	Tampa	0	0	0	55	chr10_782038	0.0532	10	0	94
chr10_1367775	Solomon	0.0086	1	0	58	chr10_782038	0.0412	8	0	97
chr10_1367775	Lupului	0.0581	5	0	43	chr10_782038	0.1474	26	1	95
chr10_1367775	Ruia	0.3725	8	15	51	chr10_782038	0.4286	56	14	98

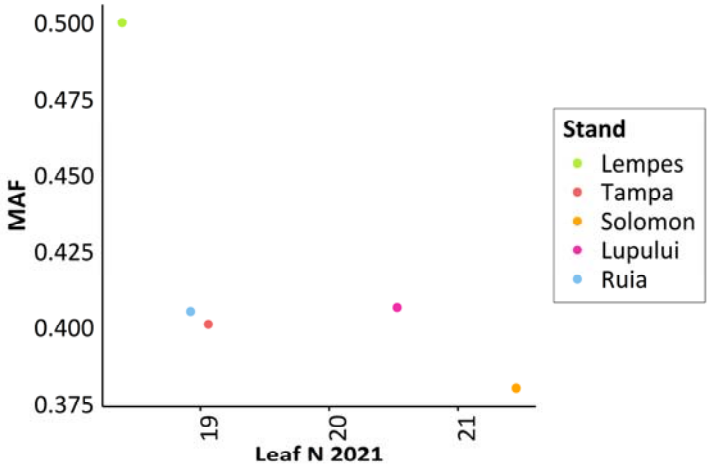
S10: Correlation between stomatal density and the minor allele frequency (MAF) of the significant marker (p ≤0.000001) on chromosome 10 at ~ 6.314 at 0.92 (p-value = 0.028) (a), ~ 7.116 at 0.9 (p-value = 0.037) (b), ~ 7.555 at 0.89 (p-value = 0.041) (c), ~ 7.7859 at 0.91 (p-value = 0.033) (d), ~ 7.82 at 0.89 (p-value = 0.045) (e), ~ 10.90916 at 0.92 (p-value = 0.029) (f), ~ 11.2335 at 0.9 (p-value = 0.039) (g), ~ 12.5471 at 0.91 (p-value = 0.033) (h), ~ 13.67776 at 0.88 (p-value = 0.047) (i).



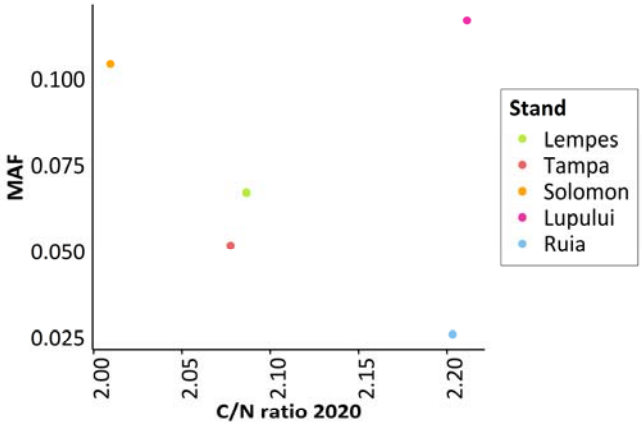
S11: Minor allele frequency (MAF) observed in the different stands at the significant markers (p ≤0.000001) associated with leaf nitrogen content and C/N ratio, the stand mean, the number of heterozygous individuals (Nr of Het), number of homozygous individuals (Nr of Hom) and the total number of genotyped individuals (Nr of total ind.).

Marker	Population	MAF	Nr of Het.	Nr of Hom.	Nr of total	Trait	Trait mean
chr2_31803661	Lempes	0.5	41	26	93	NC_2021	18.39
chr2_31803661	Tampa	0.401042	45	16	96	NC_2021	19.06
chr2_31803661	Solomon	0.380208	39	17	96	NC_2021	21.45
chr2_31803661	Lupului	0.406593	54	10	91	NC_2021	20.53
chr2_31803661	Ruia	0.405263	51	13	95	NC_2021	18.92
chr10_13027242	Lempes	0.06707	7	2	82	CNR 2020	2.09
chr10_13027242	Tampa	0.05172	7	1	87	CNR 2020	2.08
chr10_13027242	Solomon	0.10440	11	4	91	CNR 2020	2.01
chr10_13027242	Lupului	0.11702	14	4	94	CNR 2020	2.21
chr10_13027242	Ruia	0.02604	3	1	96	CNR 2020	2.20
chr11_26539169	Lempes	0.13	22	2	100	NC_2020	25.70
chr11_26539169	Tampa	0.131313	26	0	99	NC_2020	25.59
chr11_26539169	Solomon	0.086735	17	0	98	NC_2020	27.36
chr11_26539169	Lupului	0.163265	24	4	98	NC_2020	25.49
chr11_26539169	Ruia	0.176768	23	6	99	NC_2020	25.26

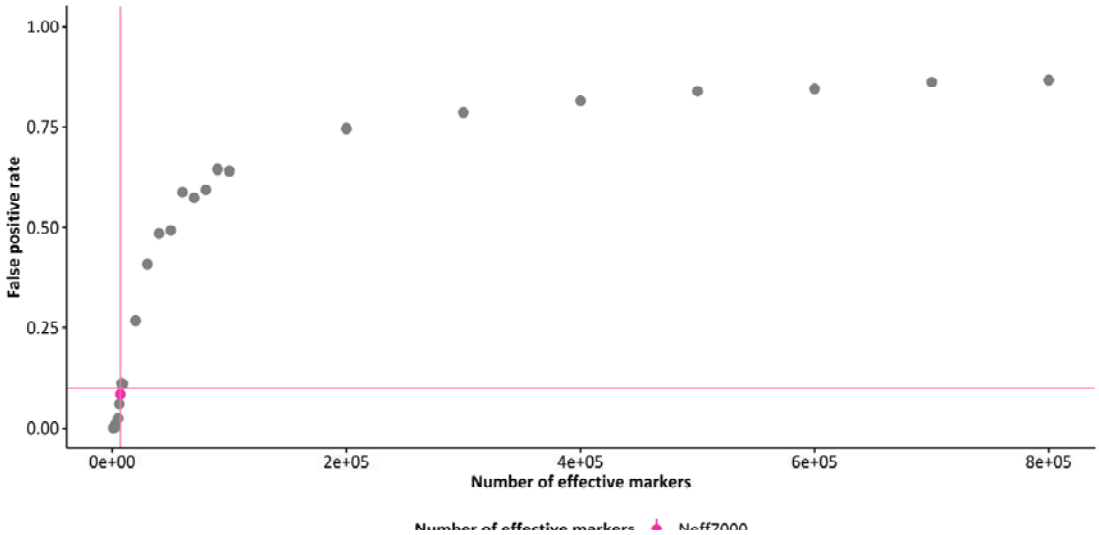
S12: Correlation between leaf nitrogen content measured in 2021 and the minor allele frequency (MAF) of the significant marker (p ≤0.000001) on chromosome 2 at ~31.804 Mb. This marker was annotated with the gene variant *Bhaga_2.g3457* with the underlying gene *XP_030960280.1*, also known as *abscisic-aldehyde oxidase-like*.



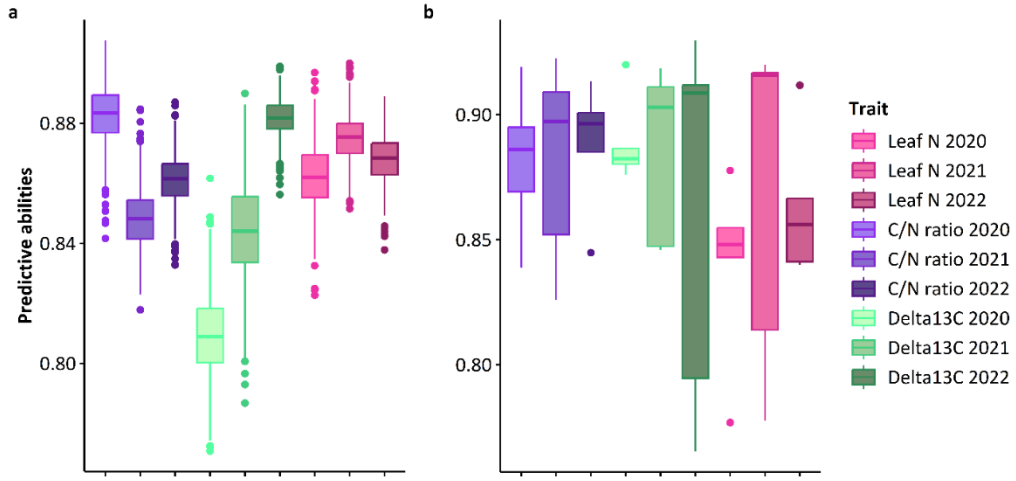
S13: Correlation between C/N ratio measured in 2020 and the minor allele frequency (MAF) of the significant marker (p ≤0.000001) on chromosome 10 at ~13.03 Mb. No annotations were found for this marker.



S14: Rate of false positive observations observed in Ghat test calculated based on the 100 randomly generated traits with 10 replications (10 replications x 100 random traits).



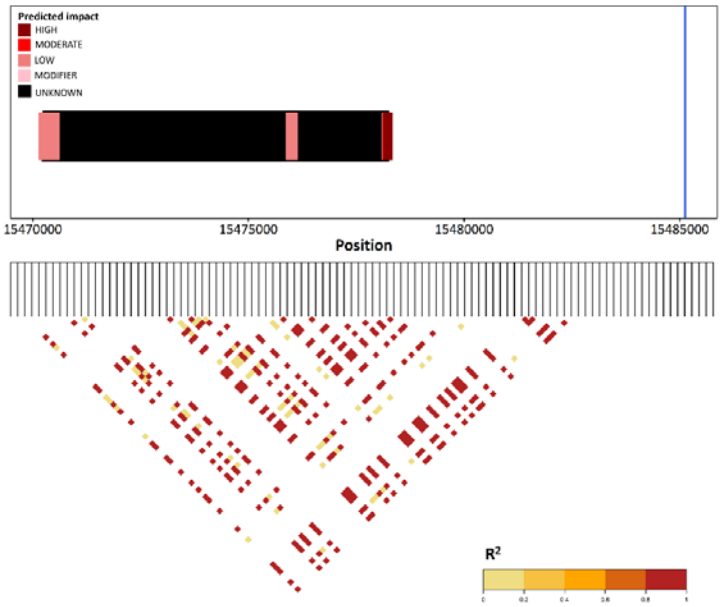
S15: Results from five-fold cross-validation with the mean predictive ability as (correlation of predicted trait estimates and observed trait measurements) with random partition of stands in test and training set (**a**) and five-fold cross-validation in which one stand was left out and the remaining stands were used for prediction (**b**) with 100 replications based on RR-BLUP for the real trait measurements.



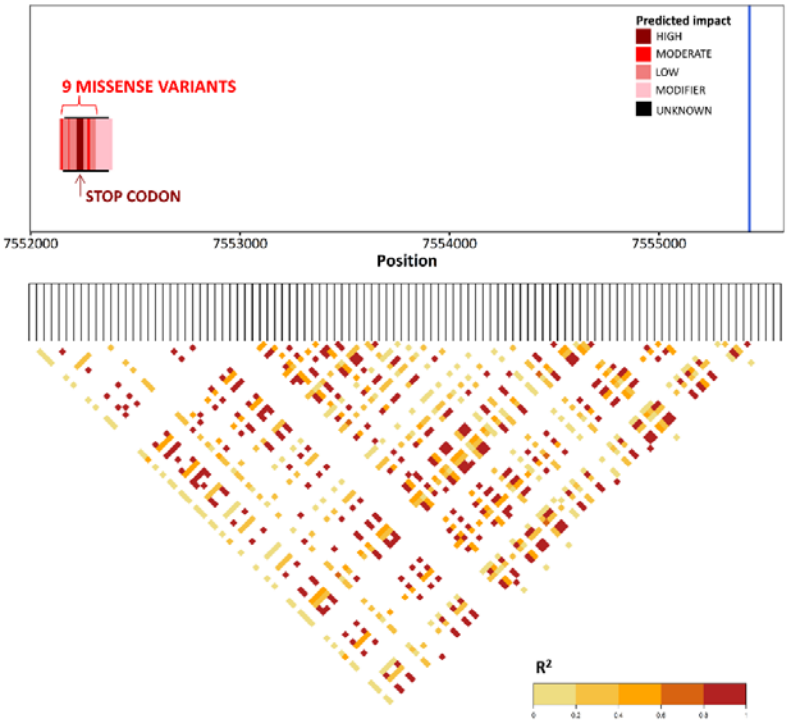
S16: Results from test with change calculated between groups (Ruia, Solomon vs Tampa, Lempes) calculated between groups with 100000 permutations and 7000 effective markers calculated based on the previous test runs for carbon isotope composition $\delta^{13}C$.

Trait	correlation	p-value
$\delta^{13}C$ (2020)	0.1017	0.0000
$\delta^{13}C$ (2021)	0.0962	0.0000
$\delta^{13}C$ (2022)	-0.017	0.1527

S17: Coding region of the gene variant *Bhaga_8.g1863* (*ABC transporter I family member 10*) and the significant marker on chromosome 8 at ~ 15.485 Mb in blue and pairwise LD heat map from 15470128 to 15500213 Mb.



S18: Coding region of the gene variant *Bhaga_10.g917* (*ALMT2*) and the significant marker on chromosome 10 at ~7.56 Mb in blue and pairwise LD heat map from 7.550433 to 7560433 Mb.

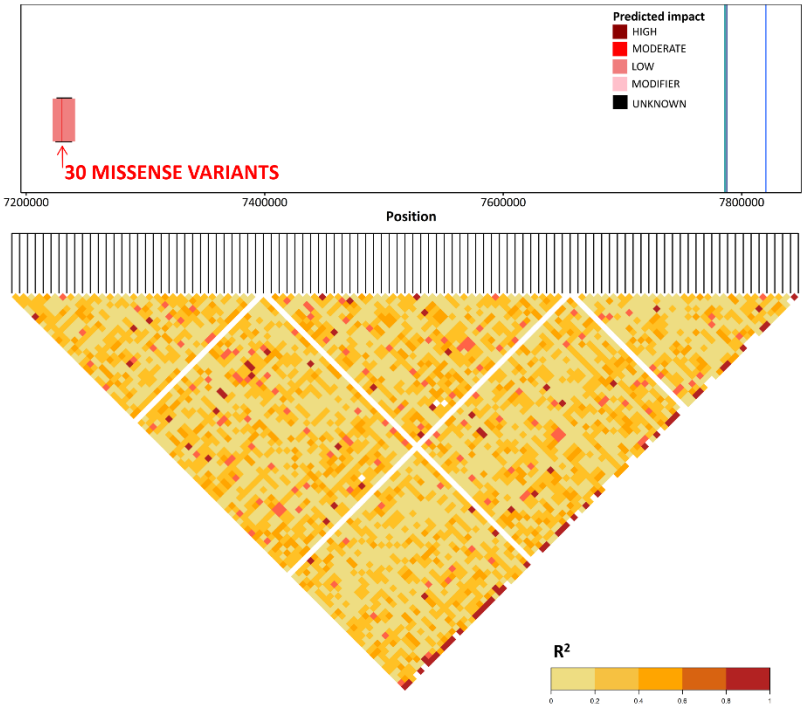


S19: Observed codon changes which lead to an amino acid (AA) replacement and a resulting change in polar bonds (hydrophobic or hydrophilic) at the underlying gene variants of markers associated with stomatal density

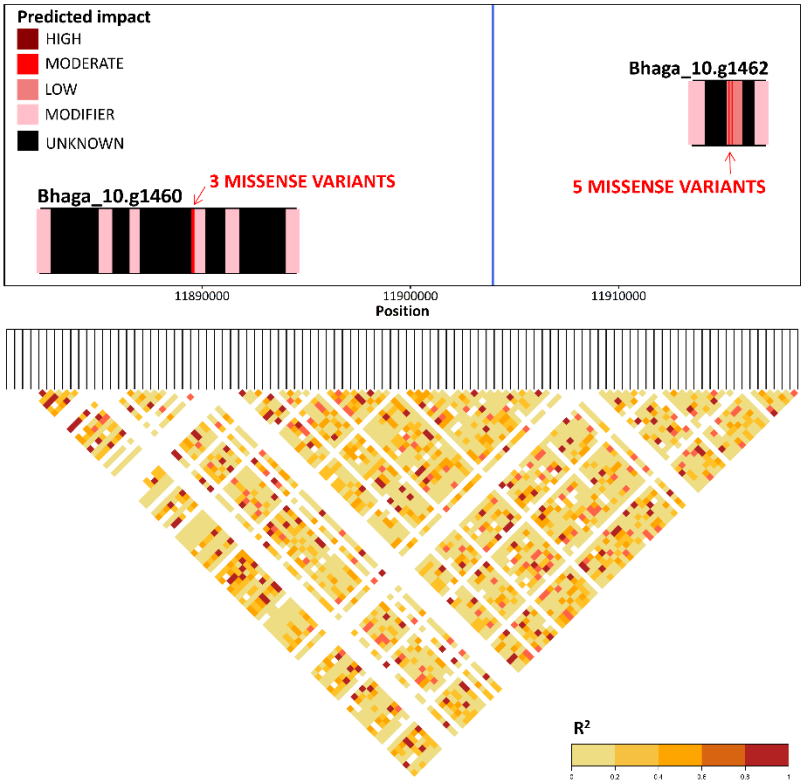
Gene variant	Variant version	Change in polar bonds	Original	Replaced by
Bhaga_8.g506	5 missense variants	Hydrophobic AA was replaced by hydrophilic AA	Pro	Ser
		Hydrophilic AA was replaced by hydrophobic AA	Glu	Asp
		Hydrophilic AA was replaced by hydrophobic AA	Arg	Lys
		Hydrophilic AA was replaced by hydrophobic AA	Val	Ala
		Hydrophobic AA was replaced by hydrophilic AA	Asn	Lys
		Hydrophilic AA was replaced by hydrophobic AA	Glu	Asp
	3 missense variants	no change in polar bonds		
Bhaga_8.g1863	1 missense variant	Hydrophilic AA was replaced by hydrophobic AA	Lys	Asn
Bhaga_9.g1680	1 missense variant	Hydrophilic AA was replaced by hydrophobic AA	Asp	His
Bhaga_9.g1682	13 missense variants	Hydrophobic AA was replaced by hydrophilic AA	Asn	Lys
		Hydrophilic AA was replaced by hydrophobic AA	Arg	Tyr
		Hydrophilic AA was replaced by hydrophobic AA	Asn	Ile
		Hydrophilic AA was replaced by hydrophobic AA	Lys	Glu
		Hydrophilic AA was replaced by hydrophobic AA	Glu	Asp
		Hydrophilic AA was replaced by hydrophobic AA	Thr	Ser
		Hydrophobic AA was replaced by hydrophilic AA	Tyr	Asp
		Hydrophilic AA was replaced by hydrophobic AA	Lys	Glu
		Hydrophilic AA was replaced by hydrophobic AA	Arg	Gly
		Hydrophobic AA was replaced by hydrophilic AA	Gly	Asp
		Hydrophobic AA was replaced by hydrophilic AA	Asn	Ser
		Hydrophilic AA was replaced by hydrophobic AA	His	Leu
		Hydrophilic AA was replaced by hydrophobic AA	Gly	Val
Bhaga_10.g591	3 missense variants	Hydrophobic AA was replaced by hydrophilic AA	Leu	His
		Hydrophobic AA was replaced by hydrophilic AA	Asn	His
		Hydrophilic AA was replaced by hydrophobic AA	Leu	Val
Bhaga_10.g752	1 missense variant	Hydrophilic AA was replaced by hydrophobic AA	Glu	Lys
Bhaga_10.g753	3 missense variants	Hydrophilic AA was replaced by hydrophobic AA	Val	Ile

		no change in polar bonds	Gln	Lys
		Hydrophobic AA was replaced by hydrophilic AA	Met	Lys
Bhaga_10.g852	1 missense variant	Hydrophilic AA was replaced by hydrophobic AA	Val	Ile
Bhaga_10.g854	7 missense variants	Hydrophilic AA was replaced by hydrophobic AA	Phe	Tyr
		Hydrophilic AA was replaced by hydrophobic AA	Leu	Phe
		Hydrophilic AA was replaced by hydrophobic AA	Ala	Val
		Hydrophobic AA was replaced by hydrophilic AA	Asn	Lys
		Hydrophilic AA was replaced by hydrophobic AA	Val	Ile
		Hydrophilic AA was replaced by hydrophobic AA	Glu	Gly
		Hydrophilic AA was replaced by hydrophobic AA	Phe	Val
Bhaga_10.g867	7 missense variants	Hydrophilic AA was replaced by hydrophobic AA	Thr	Asn
		Hydrophilic AA was replaced by hydrophobic AA	Leu	Phe
		Hydrophilic AA was replaced by hydrophobic AA	Gly	Ala
		Hydrophilic AA was replaced by hydrophobic AA	His	Leu
		Hydrophilic AA was replaced by hydrophobic AA	Ser	Trp
		Hydrophobic AA was replaced by hydrophilic AA	Asn	His
		Hydrophobic AA was replaced by hydrophilic AA	Asn	Lys
Bhaga_10.g917	9 missense variants	Hydrophilic AA was replaced by hydrophobic AA	Phe	Leu
		Hydrophilic AA was replaced by hydrophobic AA	Asp	Ala
		Hydrophilic AA was replaced by hydrophobic AA	Ser	Cys
		Hydrophilic AA was replaced by hydrophobic AA	Leu	Phe
		Hydrophobic AA was replaced by hydrophilic AA	Ile	Ser
		Hydrophobic AA was replaced by hydrophilic AA	Gly	Glu
		Hydrophobic AA was replaced by hydrophilic AA	Gly	Arg
		Hydrophilic AA was replaced by hydrophobic AA	Ser	Leu
		Hydrophilic AA was replaced by hydrophobic AA	Ser	Asn
Bhaga_10.g957	5 missense variants	Hydrophilic AA was replaced by hydrophobic AA	Glu	Lys
		Hydrophobic AA was replaced by hydrophilic AA	Asn	Lys
		Hydrophobic AA was replaced by hydrophilic AA	Pro	Thr
		Hydrophilic AA was replaced by hydrophobic AA	Ala	Val
		Hydrophilic AA was replaced by hydrophobic AA	Arg	Gln
	25 missense variants	no change in polar bonds		
Bhaga_10.g1460	2 missense variants	Hydrophilic AA was replaced by hydrophobic AA	Gly	Ala
		Hydrophilic AA was replaced by hydrophobic AA	Ile	Val
	1 missense variant	no change in polar bonds		
Bhaga_10.g1462	5 missense variants	Hydrophobic AA was replaced by hydrophilic AA	Ala	Ser
		Hydrophilic AA was replaced by hydrophobic AA	Thr	Lys
		Hydrophobic AA was replaced by hydrophilic AA	Gly	Arg
		Hydrophilic AA was replaced by hydrophobic AA	Asp	Asn
		Hydrophilic AA was replaced by hydrophobic AA	Asp	Glu
Bhaga_10.g1543	3 missense variants	Hydrophobic AA was replaced by hydrophilic AA	Tyr	His
		Hydrophobic AA was replaced by hydrophilic AA	Phe	Cys
		Hydrophilic AA was replaced by hydrophobic AA	Ser	Thr
Bhaga_10.g1652	1 missense variant	Hydrophilic AA was replaced by hydrophobic AA	Glu	Lys
Bhaga_10.g1674	5 missense variants	Hydrophilic AA was replaced by hydrophobic AA	Ile	Leu
		Hydrophilic AA was replaced by hydrophobic AA	Val	Ala
		Hydrophilic AA was replaced by hydrophobic AA	Arg	Pro
		Hydrophilic AA was replaced by hydrophobic AA	Arg	Gly
		Hydrophobic AA was replaced by hydrophilic AA	Leu	His

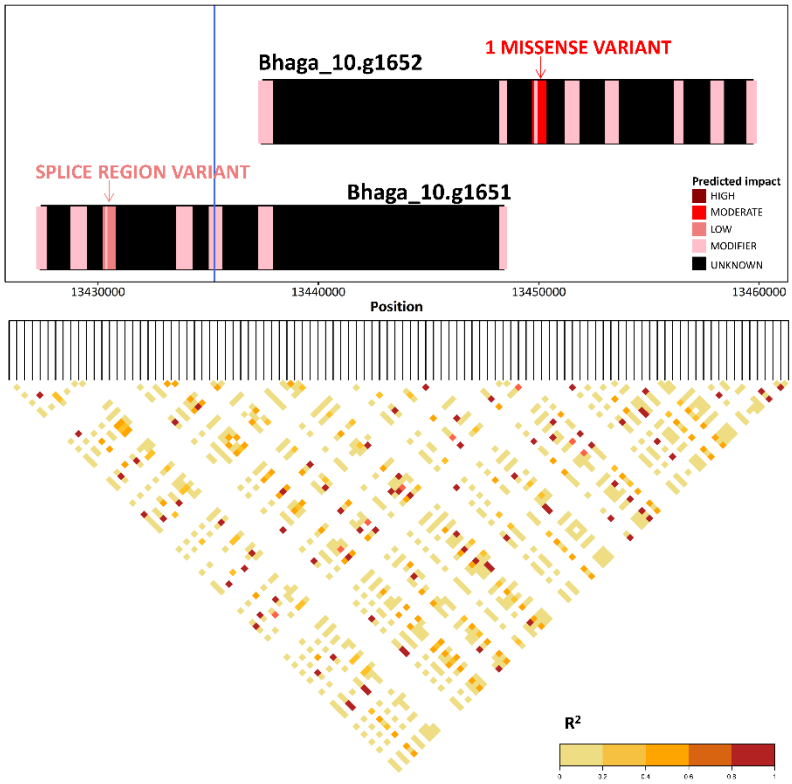
S20: Coding region of the gene variant *Bhaga_10.g957* (*MYB35-like*) and the five significant markers on chromosome 10 at ~7.78 to 7.82 Mb (7785927; 7785978; 7787468; 7820271; 7820389) Mb in blue and pairwise LD heat map from 7.765927 to 7.840389 Mb.



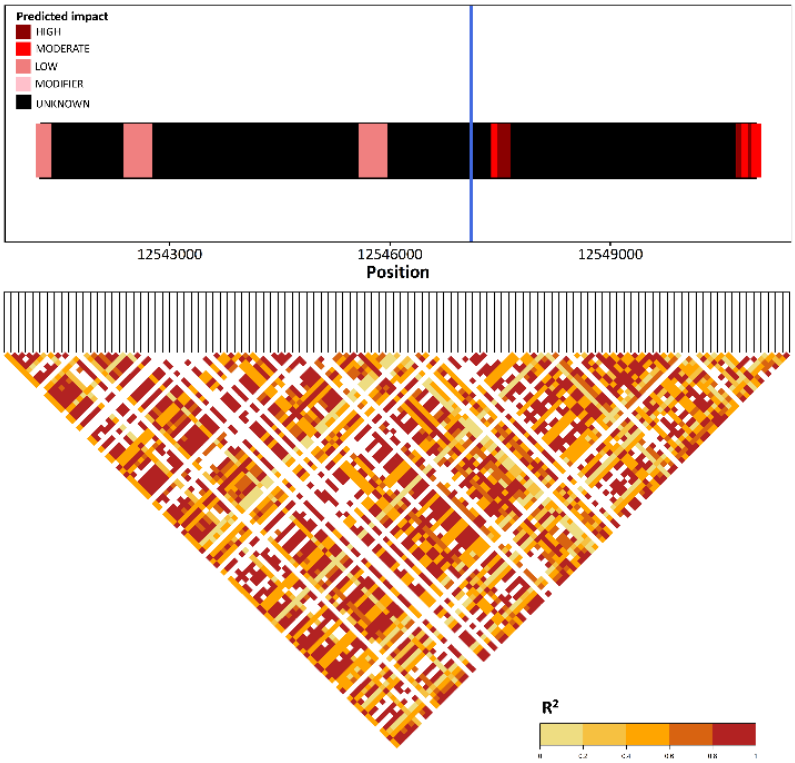
S21: Coding region of the gene variants *Bhaga_10.g1460 (ClpS1)*, *Bhaga_10.g1462* (*ASPG1*) and the significant marker on chromosome 10 at ~11.9 Mb in blue and pairwise LD heat map from 11.888957 to 11.918957 Mb.



S22: Coding region of the gene variant *Bhaga_10.g1651 (BONZAI1)*, *Bhaga_10.g1652* (uncharacterized) and the significant marker on chromosome 10 at ~ 13.44 Mb in blue and pairwise LD heat map from 13420289 to 13440289 Mb.



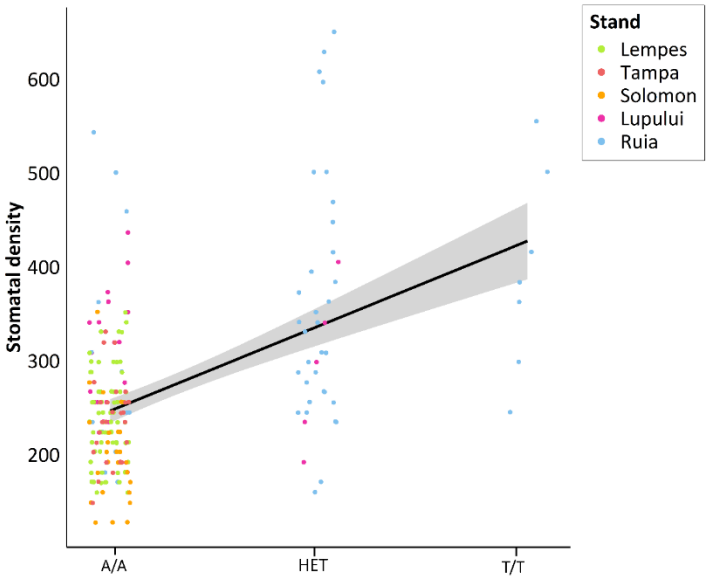
S23: Coding region of the gene variant *Bhaga_10.g1543* (*B10-like*) and the significant marker on chromosome 10 at ~ 12.541 Mb in blue and pairwise LD heat map from 12545109 to 12549109 Mb.

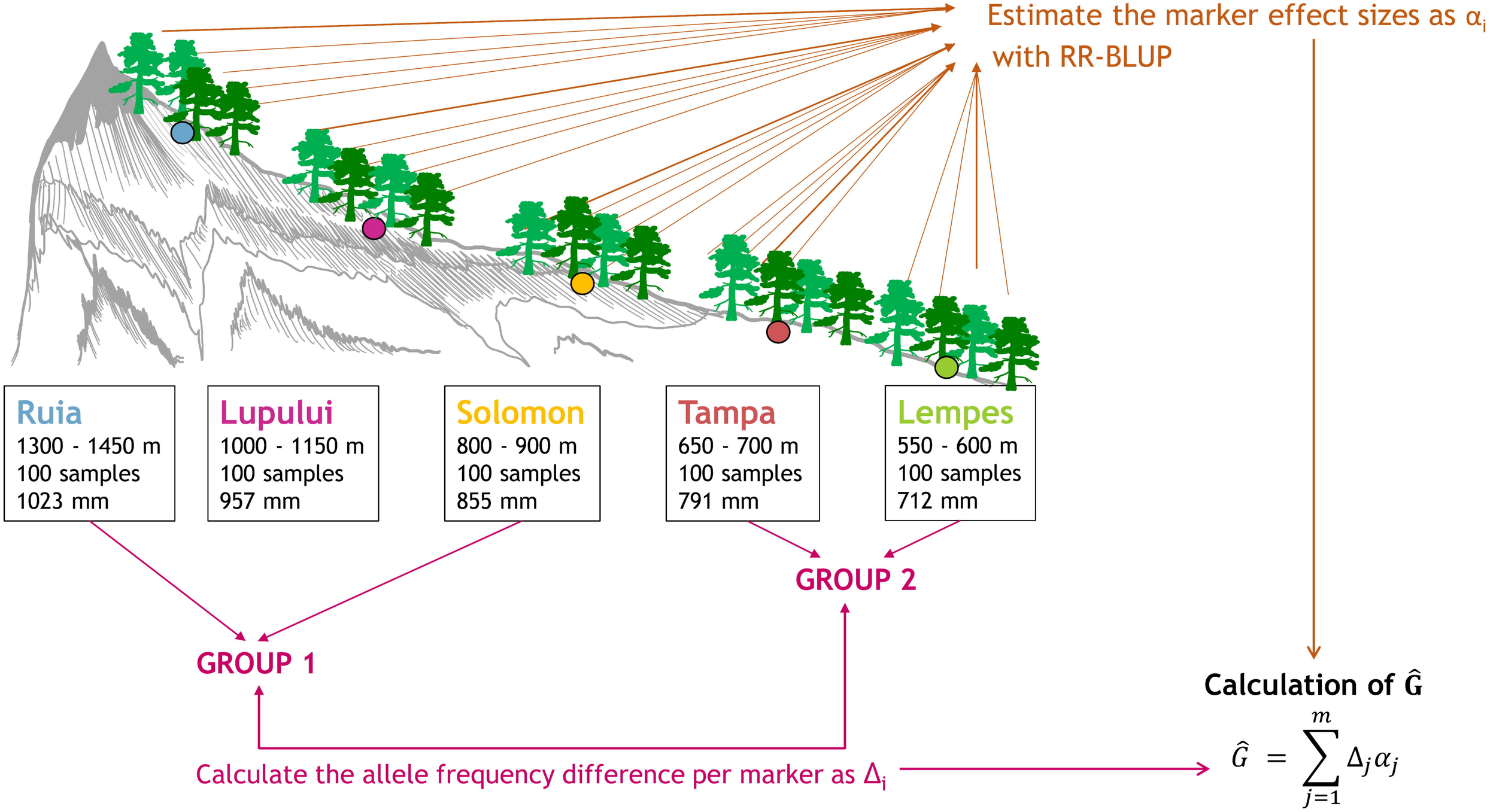


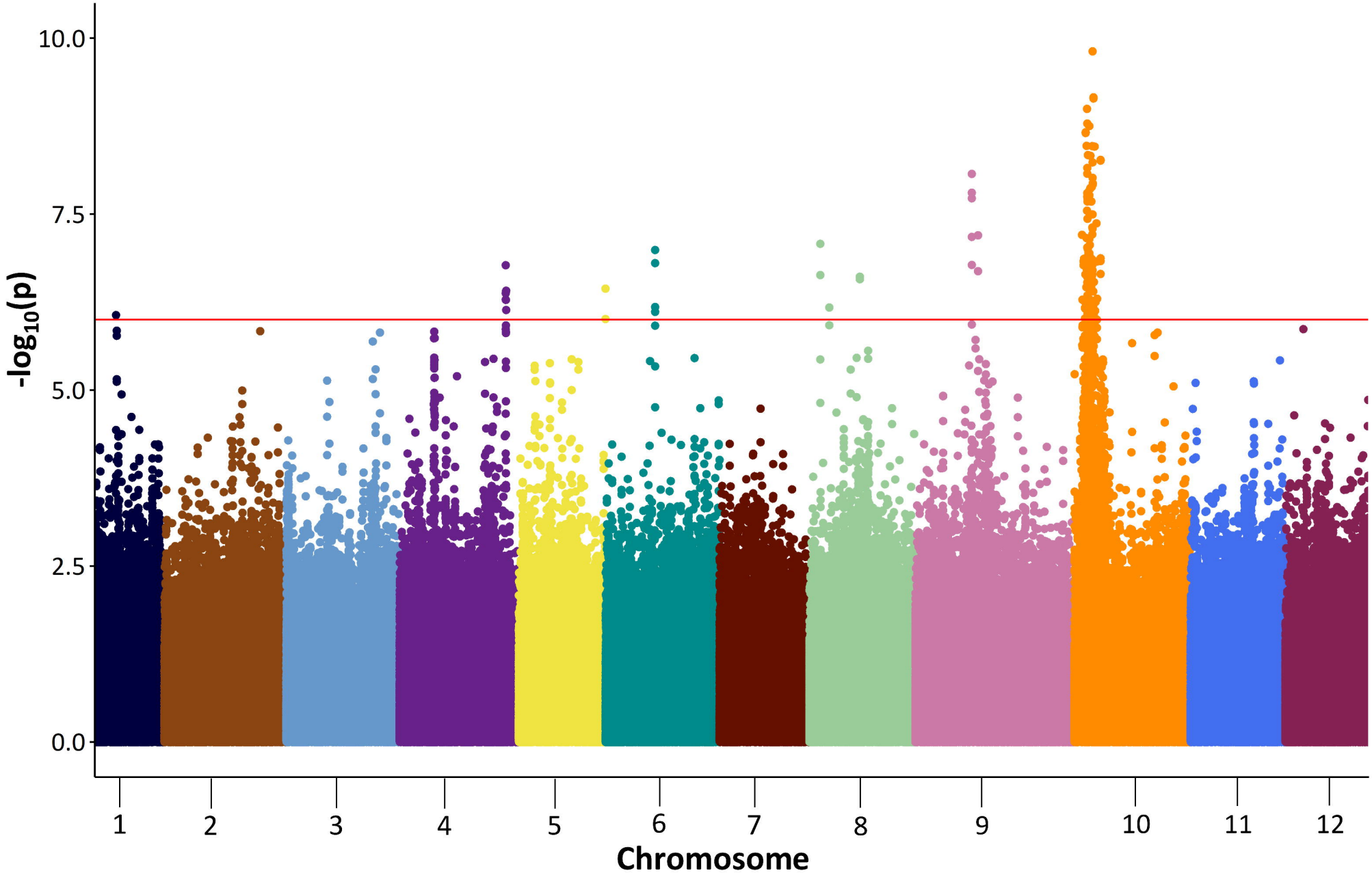
S24: List of other genes close to markers associated with stomatal density but without a direct impact on stomatal density based on the literature review

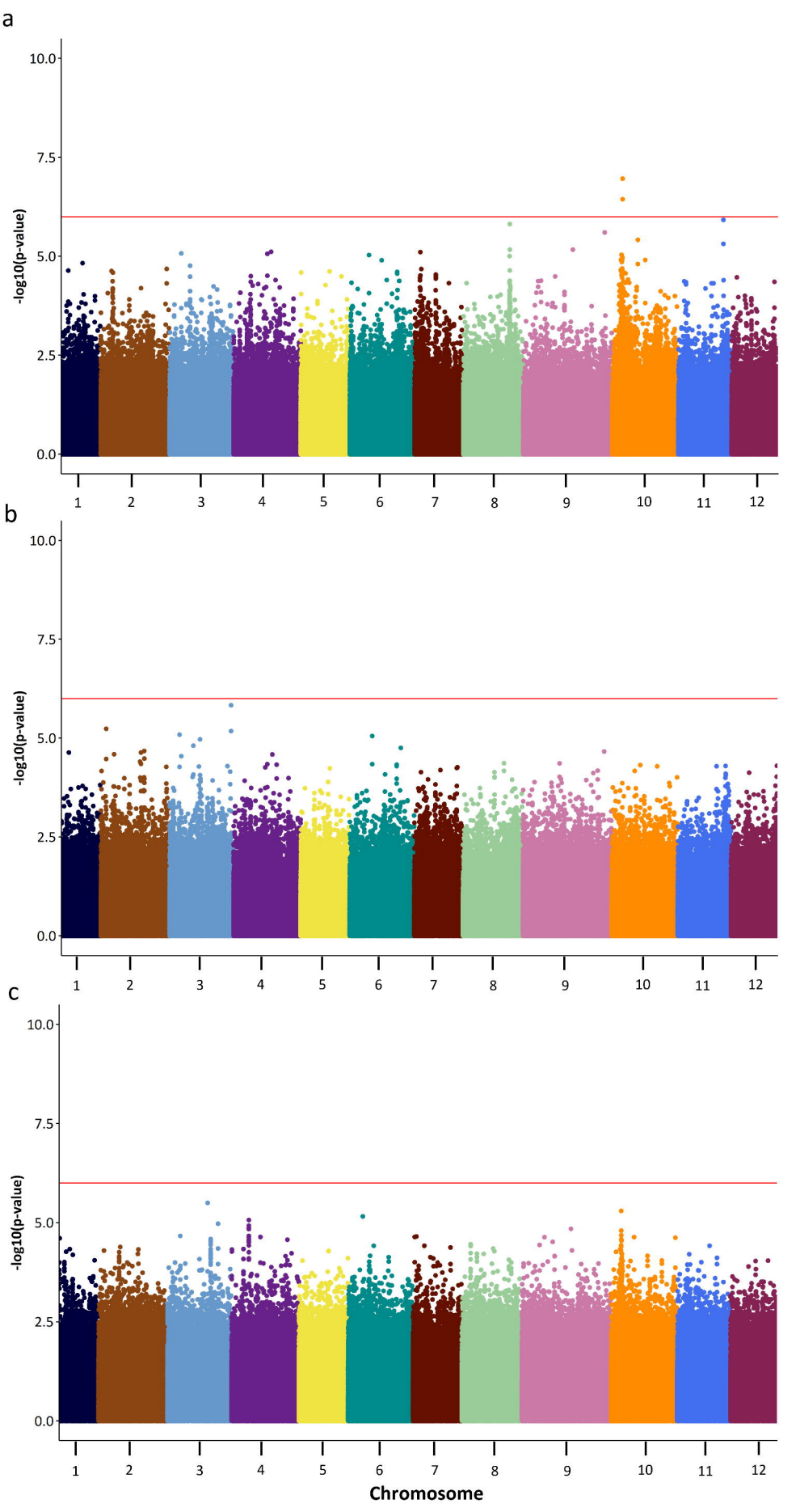
Gene variant name	Gene	Gene description based on the literature review
Bhaga_8.g506	**TFbHLH117**	*TFbHLH117* is a transcription factor of basic-helix-loop helices (bHLH), which is one of the largest families of transcription factors found in plants (Gao and Dubos 2024).
Bhaga_10.g752	**Histone deacetylase HDT1-like**	*Histone deacetylase HDT1-like* appears to be involved in stem vascular development (Zhang et al. 2019b)
Bhaga_10.g854	**Germin-like protein 9-3**	Germin-like protein 9-3 was observed to be involved in a broad-spectrum of disease resistance in rice (Manosalva et al. 2009). Gangadhar et al. (2021) reports that improved heat stress tolerance was observed in potato due to germin-like proteins.
Bhaga_10.g867	**Capsanthin synthase chromoplastic-like**	Capsanthin synthase chromoplastic-like is mainly studied and observed in species of the *Capsicumin* genus like carrot (Deng et al. 2024), paprika, pepper or chili pepper (Piano et al. 2019), but also found in some species of *Lilium leichtinii* (lily) (Valadon and Mummery, 1976) and asparagus (Deli et al. 2000).
Bhaga_10.g873	**mRNA-decapping enzyme-like protein**	The gene mRNA-decapping enzyme-like protein regulates mRNA decapping (Vidya and Duchaine 2022). Decapping of the mRNA means the removal of the m7G cap, which leads quickly to mRNA degradation (Vidya and Duchaine 2022).
Bhaga_10.g1543	**Glycosyltransferase BC10-like**	Glycosyltransferase BC10-like belongs to the family of glycosyl transferases. This gene was only studied in rice (*Oryza sativa* L.) (Zhang et al. 2016) and required for regulation of cellulose biosynthesis in the cell wall (Zhang et al. 2016).
Bhaga_10.g1562	**Kinesin-like protein NACK1**	Kinesin-like protein NACK1, which is a key regulator of plant cytokinesis (Sasabe et al. 2011). Kinesin-like protein NACK1 forms with NPK1 a complex (Nishihama et al. 2002). This complex stimulates kinase activity and plays an essential role in intracellular events that lead to cytokinesis (Nishihama et al. 2002).
Bhaga_10.g1674	**poly(A) polymerase I-like gene**	The poly(A) polymerase I-like gene is responsible 3'polyadenylation which leads to rapid degradation of 3' to 5' exonucleases (Feng and Cohen 2000). This gene was so far only studied in E. coli (Feng and Cohen 2000).

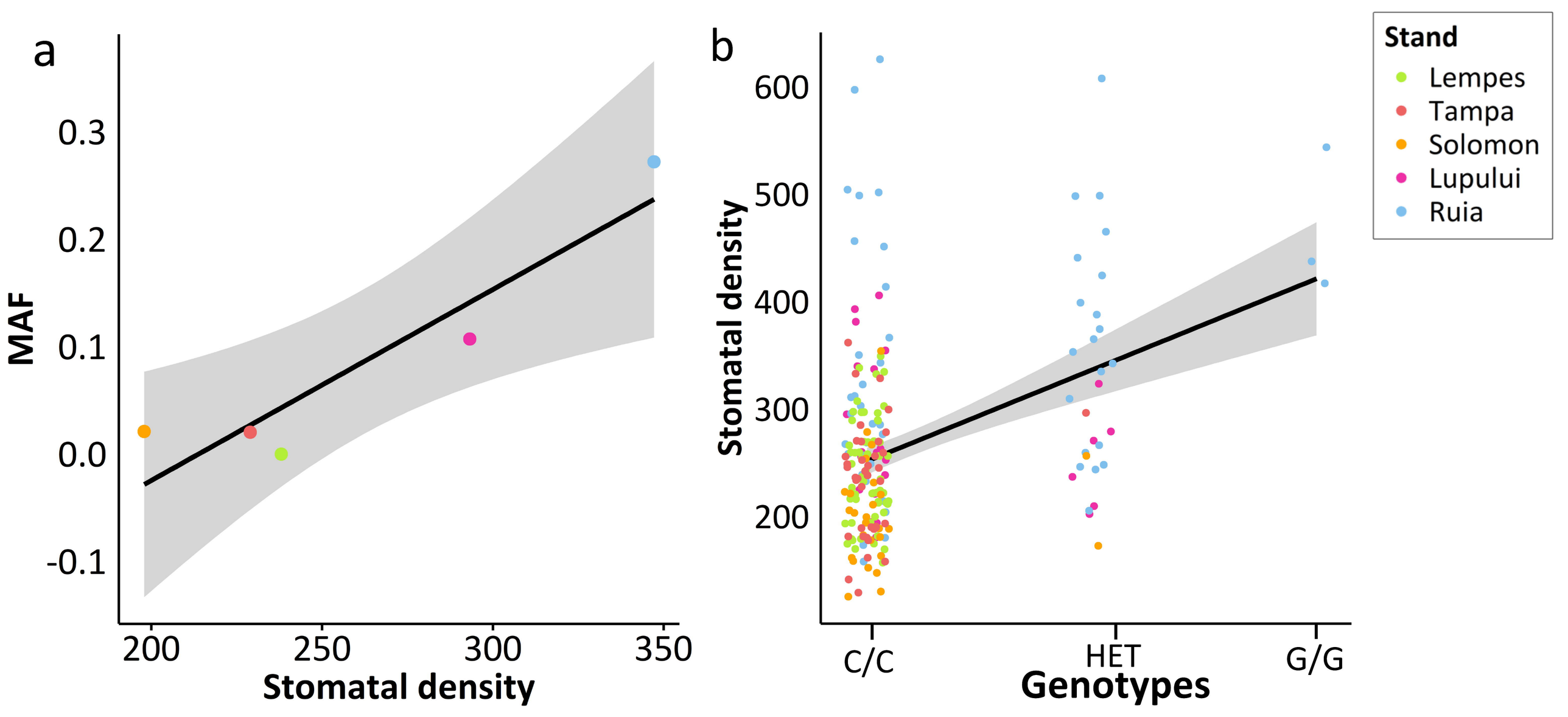
S25: Correlation between stomatal density and observed genotypes at the significant marker ($p \leq 0.000001$) on chromosome 10 at ~7.82 Mb. The coefficient of determination R^2 was 0.2353 ($p \sim 0$).

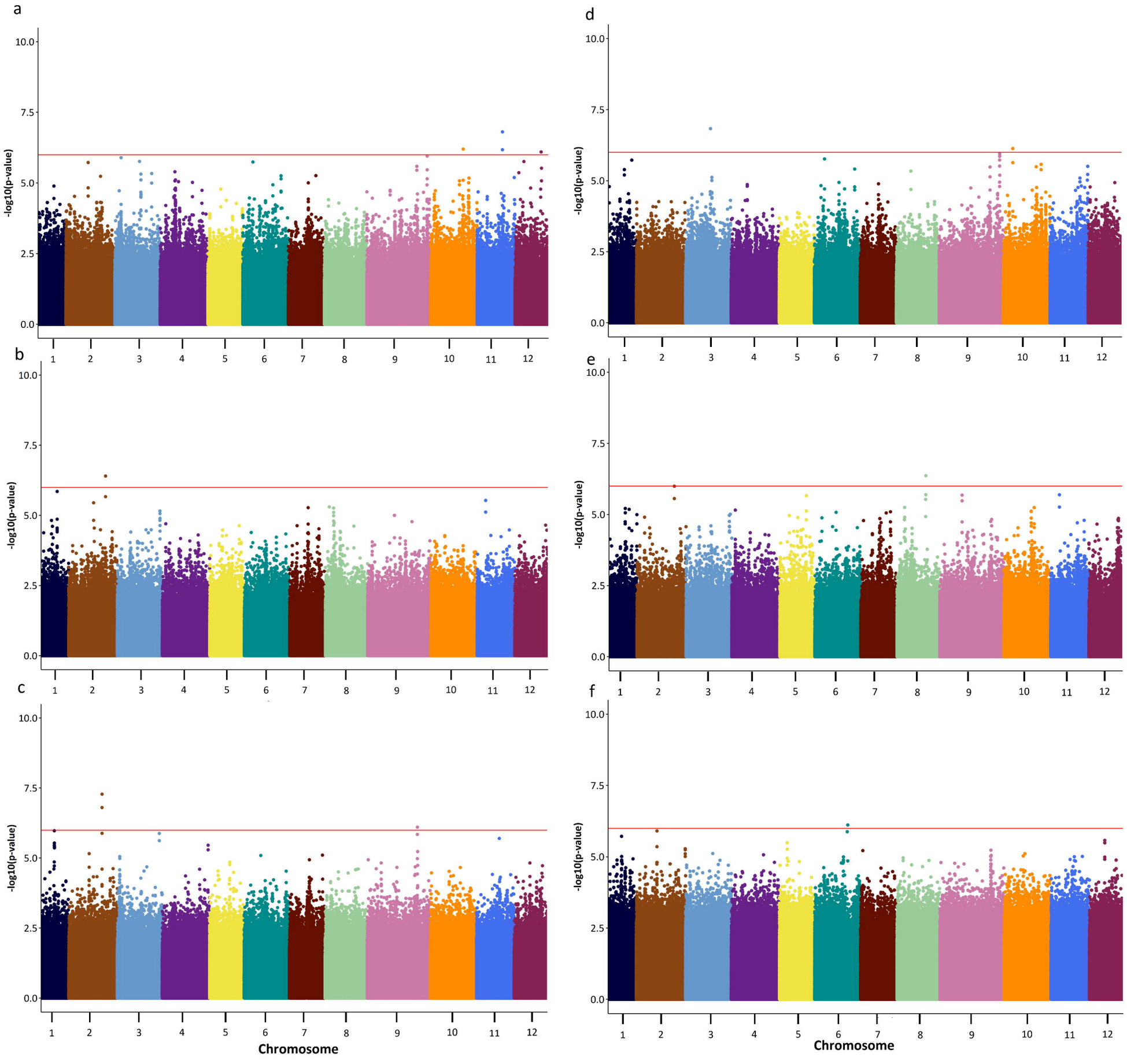


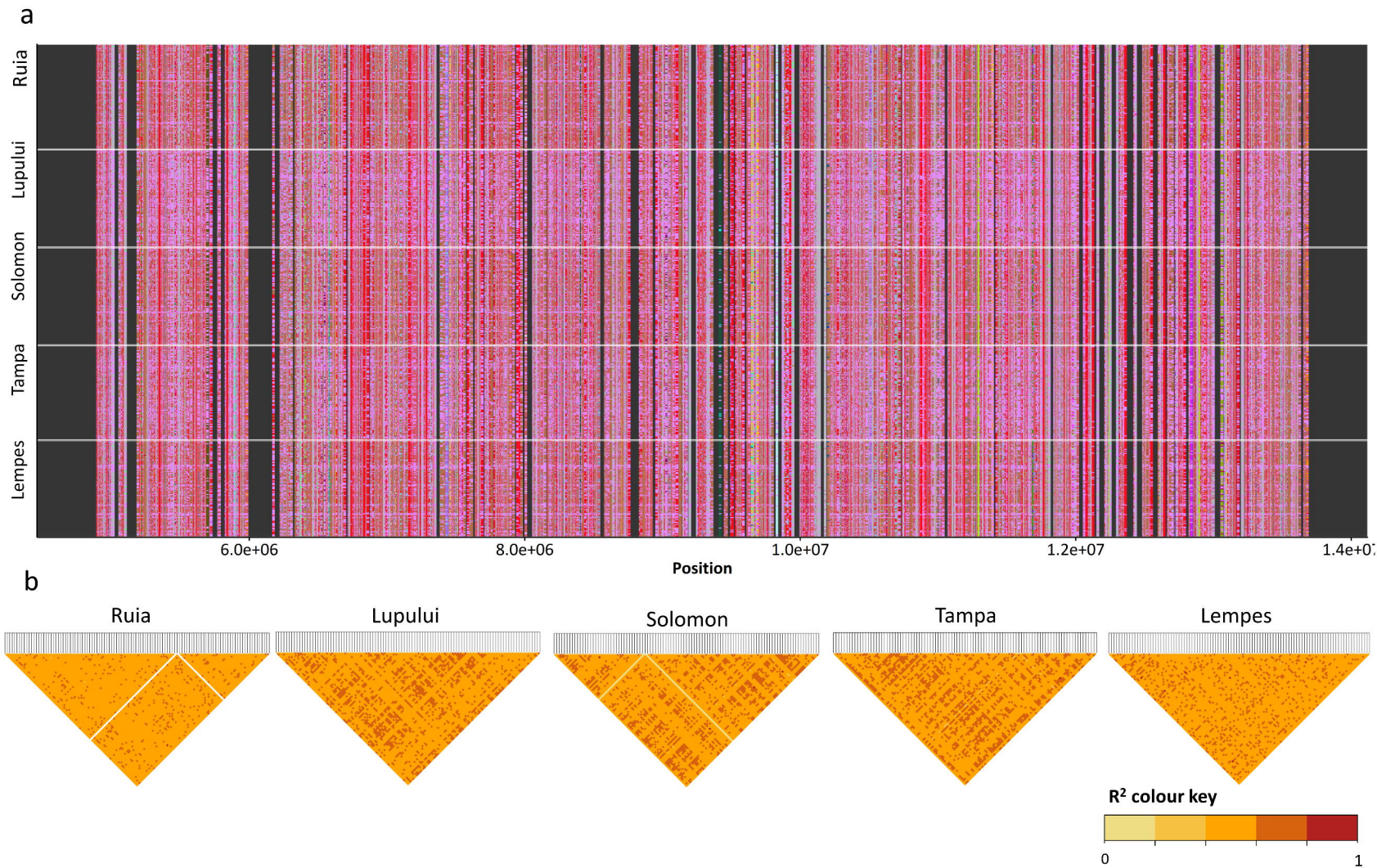

Estimate the marker effect sizes as α_i with RR-BLUP
Ruia
1300 - 1450 m
100 samples
1023 mm
Lupului
1000 - 1150 m
100 samples
957 mm
Solomon
800 - 900 m
100 samples
855 mm
Tampa
650 - 700 m
100 samples
791 mm
Lempes
550 - 600 m
100 samples
712 mm
GROUP 2
GROUP 1
Calculation of $\hat{G}$
$\hat{G} = \sum_{j=1}^{m} \Delta_j \alpha_j$
Calculate the allele frequency difference per marker as Δ_i

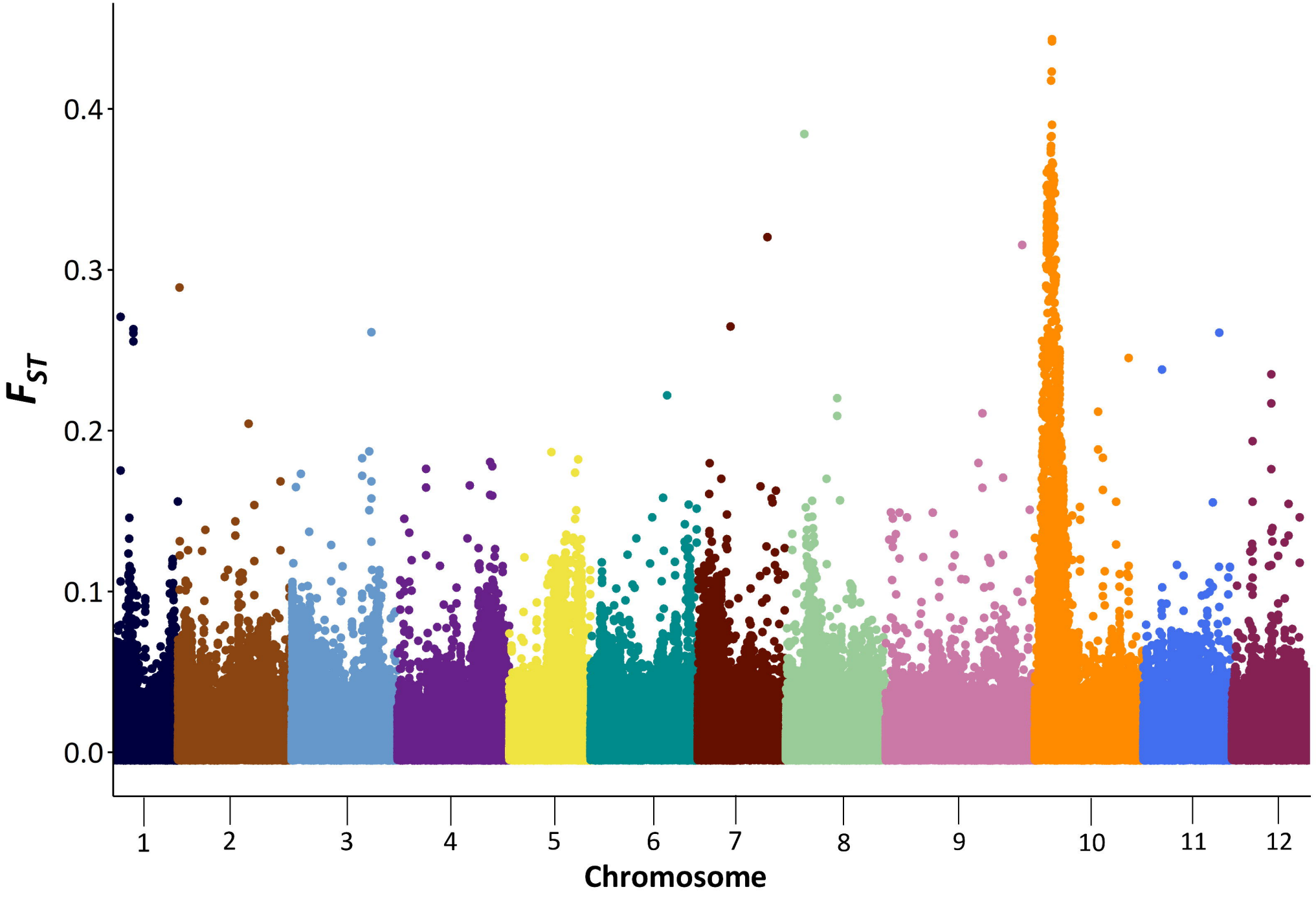

10.0
7.5
5.0
2.5
0.0
-log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome


a
b
c
-log10(p-value)
10.0
7.5
5.0
2.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome


a
MAF
0.3
0.2
0.1
0.0
-0.1
200
250
300
350
Stomatal density
b
Stomatal density
600
500
400
300
200
C/C
HET
G/G
Genotypes
Stand
Lempes
Tampa
Solomon
Lupului
Ruia


a
b
c
d
e
f
-log10(p-value)
10.0
7.5
5.0
2.5
0.0
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome


a
Ruia
Lupului
Solomon
Tampa
Lempes
6.0e+06
8.0e+06
1.0e+07
1.2e+07
1.4e+07
Position
b
Ruia
Lupului
Solomon
Tampa
Lempes
R² colour key
0
1


0.4
0.3
0.2
0.1
0.0
F_{ST}
1
2
3
4
5
6
7
8
9
10
11
12
Chromosome