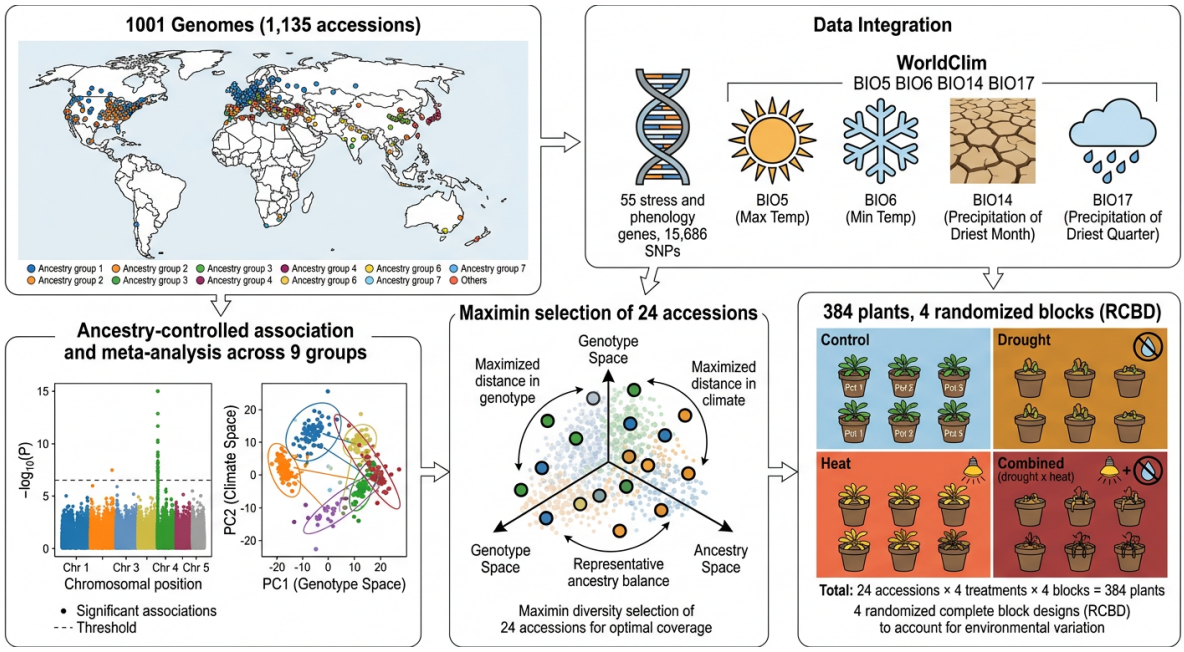
Climate-Informed, Ancestry-Controlled Nomination of 24 *Arabidopsis thaliana* Accessions for a Factorial Drought-by-Heat Common-Garden Experiment

A Technical Report on Panel Design, Locus-Level Hypotheses,
Population-Structure Diagnostics, and Experimental Layout



Graphical abstract. From the 1001 Genomes (GMI–MPI v3.1) 1,135-accession release and WorldClim v2.1 climate normals, variants in a preregistered stress and phenology gene panel are extracted, associations with four bioclimatic variables are estimated within ancestry groups and meta-analyzed, and 24 accessions are nominated by maximin diversity across genotype, provenance climate, and ancestry for a 2 × 2 factorial (drought × heat) randomized complete block experiment.

Reproducible workflow, fixed random seed 42. Data sources: 1001 Genomes Consortium (GMI–MPI v3.1); WorldClim v2.1 (1970–2000 normals).

July 14, 2026

Abstract

Designing common-garden experiments that both span the adaptive diversity of a species and remain statistically tractable requires principled accession selection rather than convenience sampling. Here we document a fully reproducible pipeline that nominates 24 *Arabidopsis thaliana* accessions for a factorial drought-by-heat experiment from the 1001 Genomes GMI–MPI v3.1 release (1,135 natural inbred lines) and WorldClim v2.1 climate normals. We restricted variant extraction to a preregistered panel of 55 drought-, heat-, and phenology-related genes (±10 kb windows), yielding 15,686 quality-controlled SNPs, and annotated every accession with four bioclimatic variables at its collection coordinates: maximum temperature of the warmest month (BIO5, a heat proxy), minimum temperature of the coldest month (BIO6), and precipitation of the driest month and driest quarter (BIO14, BIO17; drought proxies). Population structure was diagnosed by principal-component analysis on an LD-pruned marker set (2,134 SNPs; PC1–PC10 explaining 24.6% of variance), recovering the known continental structure of the panel and the divergent Relict lineage. Genotype–climate associations were estimated *within* each of nine ancestry groups ($N \geq 50$), adjusting for the top five PCs via Frisch–Waugh–Lovell residualization, and combined across groups by inverse-variance-weighted meta-analysis with Cochran's Q/I^2 heterogeneity statistics. This two-tier structure control reduced the genomic inflation factor 6–15× relative to a naive pooled model. Locus-level hypotheses with quantified uncertainty implicated canonical adaptation genes—*HSFA2* and *FT* for heat (BIO5), *FLC*/*ABF2* for cold (BIO6), and *NCED3*, *ABI3*, and *HSP101* for aridity (BIO14/BIO17). Rather than ranking by association strength alone, we selected 24 accessions by greedy maximin optimization of a combined genotype/climate/ancestry distance, capturing all nine genetic groups (including two Relicts) and the full climatic range of the pool. Finally, we specify a 2 × 2 factorial (watering × temperature) randomized complete block design—4 blocks, 384 plants, 16 replicates per accession—with a validated layout and a mixed-effects analysis plan. We deliver the accession panel, the population-structure diagnostics, the locus-level hypotheses, and the ready-to-execute experimental design.

Contents

1 Introduction **3**

1.1 Drought and heat as interacting constraints on plant fitness . . . 3

1.2 *Arabidopsis thaliana* as a model for climate adaptation genetics . . . 3

1.3 Genomic and climatic resources . . . 3

1.4 From association to design: why maximin diversity . . . 4

2 Methods **4**

2.1 Accession metadata and climate extraction . . . 4

2.2 Preregistered gene panel and variant extraction . . . 4

2.3 Population-structure diagnostics . . . 5

2.4 Ancestry-controlled association and meta-analysis . . . 5

2.5 Maximin diversity selection . . . 6

2.6 Factorial drought × heat design . . . 6

3 Results **6**

3.1 Climatic breadth of the source panel . . . 6

3.2 Gene panel and variant catalog . . . 7

3.3 Population-structure diagnostics . . . 8

3.4 Ancestry-controlled associations and locus-level hypotheses . . . 10

3.5 The 24 nominated accessions 13
3.6 Factorial drought × heat layout 15

4 Discussion **17**
4.1 Statistical power and the analysis plan 17
4.2 Biological interpretation of the locus-level hypotheses 18
4.3 Why maximin rather than top-hit selection 18
4.4 Limitations 18
4.5 Conclusion 19

Data and code availability **19**

1 Introduction

1.1 Drought and heat as interacting constraints on plant fitness

Drought and high temperature are among the most consequential abiotic constraints on terrestrial plant productivity and species distributions, and they are projected to intensify and co-occur more frequently under continued climate change [1, 2]. Critically, the two stresses do not simply add: plants exposed to simultaneous water deficit and heat mount a molecular response that is distinct from—and often antagonistic to—the response to either stress alone. In a landmark physiological study, Rizhsky et al. [3] showed that *Arabidopsis thaliana* plants under combined drought and heat close their stomata (a drought response) yet must dissipate heat load (which transpiration would normally accomplish), leading to leaf temperatures higher than under heat alone and to a transcriptome dominated by neither single-stress program. This "stress combination" has since been recognized as a general principle, whereby unique combinatorial responses cannot be extrapolated from single-factor experiments [2, 4]. Understanding the genetic architecture of tolerance to combined drought × heat therefore requires experimental designs that explicitly cross the two factors and that sample the natural allelic variation shaped by these pressures.

1.2 *Arabidopsis thaliana* as a model for climate adaptation genetics

A. thaliana is an outstanding system for dissecting the genetic basis of climate adaptation. It is a predominantly selfing annual with a compact, well-annotated genome and a global distribution spanning Mediterranean, temperate, boreal, and montane climates, so that natural accessions constitute a natural experiment in local adaptation [5, 6]. The molecular physiology of its stress responses is exceptionally well characterized. Water deficit is sensed and transduced largely through the phytohormone abscisic acid (ABA): the rate-limiting biosynthetic enzyme NCED3 controls ABA accumulation under dehydration [7], ABA is perceived by PYR/PYL receptors that inhibit PP2C phosphatases and release SnRK2 kinases, and downstream bZIP transcription factors of the AREB/ABF family and AP2/ERF factors such as the DREB2 proteins reprogram the transcriptome [8–11]. Heat stress is governed by a conserved network of heat-shock transcription factors (HSFs) and molecular chaperones: HSFA2 sustains acquired thermotolerance [12], the master regulators HSFA1a/b/d initiate the response, and the ClpB/HSP100 chaperone HSP101 is essential for survival of severe heat [13–15]. Because temperature also gates the transition to flowering—through the vernalization repressor FLC, the florigen FT, and thermosensory regulators such as FLM and SVP—phenology genes are integral to the climate-adaptation phenotype [16]. These mechanistic anchors motivate a candidate-gene panel that jointly covers drought, heat, and phenology.

1.3 Genomic and climatic resources

Two community resources make climate-informed accession selection feasible. First, the 1001 Genomes Consortium sequenced 1,135 natural inbred lines and released a whole-genome variant catalog (the GMI–MPI v3.1 imputed SNP matrix) together with ADMIXTURE-based ancestry assignments that partition the panel into nine geographic groups plus an "Admixed" class, and that isolate a deeply divergent **Relict** lineage ($n = 25$) surviving in glacial refugia of Iberia and North Africa [17–19]; complementary epigenomic profiling of the same collection further documents heritable variation layered on this structure [20]. Second, WorldClim provides high-resolution interpolated climate surfaces, including 19 "bioclimatic" variables that summarize temperature and precipitation seasonality; version 2.1 improves accuracy relative to the original release [21, 22]. Georeferencing each accession against WorldClim yields a provenance climate vector that serves both as a phenotype for genotype–environment association and as a selection axis.

1.4 From association to design: why maximin diversity

Beginning with genome-wide association studies of large accession panels [23], a substantial body of work has mapped climate-associated alleles in *A. thaliana* and demonstrated that such alleles predict fitness in the field [24–28]. However, a persistent methodological hazard is that genotype and environment are both structured by shared ancestry, so unadjusted associations are inflated by population stratification [29–31]. Equally important for *experimental design*, selecting accessions by association strength alone tends to over-represent a few haplotypes and geographic regions, collapsing the very diversity a common garden is meant to interrogate and sacrificing power for genotype-by-environment (G×E) inference [32, 33]. We therefore separate the two goals: we *estimate* ancestry-controlled genotype–climate associations to generate falsifiable locus-level hypotheses, but we *select* accessions to maximize joint diversity across genotype, provenance climate, and ancestry, using a maximin (max–min distance) criterion. This report documents the complete pipeline and delivers four products requested for the drought-by-heat study: (i) the 24-accession panel; (ii) locus-level hypotheses with uncertainty; (iii) population-structure diagnostics; and (iv) a factorial drought × heat experiment with matched controls.

2 Methods

2.1 Accession metadata and climate extraction

The 1001 Genomes accession table (identifier, name, CS number, country, latitude/longitude, collector, sequencing centre, and ADMIXTURE group) was obtained for all 1,135 accessions; group labels follow the nine-group + Admixed scheme of Alonso-Blanco et al. [17], including the 25-member Relict lineage. WorldClim v2.1 1970–2000 climate normals (2.5-arc-minute resolution) were downloaded, and only the four bioclimatic layers relevant to the drought-by-heat design were retained (Table 1). Each accession's four bioclimatic values were sampled from the rasters at its collection coordinates with a nearest-valid-pixel fallback (search radius ≤ 10 pixels) for coastal or masked points. Extraction quality was confirmed by internal consistency checks that must hold by definition: BIO5 $\geq$ BIO6 for 100% of georeferenced accessions, and BIO17 (driest quarter) $\geq$ BIO14 (driest month) for 100%. Of 1,135 accessions, 1,131 received complete climate vectors; four laboratory/reference strains lacking wild collection coordinates were flagged and excluded from all climate-based analyses.

Table 1: The four WorldClim v2.1 bioclimatic variables used as provenance-climate axes. BIO5 is the heat proxy; BIO14 and BIO17 are drought (aridity) proxies; BIO6 captures winter cold.

Variable	Definition	Role in design	Pool range
BIO5	Max. temperature of warmest month	Heat stress proxy	7.9–36.0 °C
BIO6	Min. temperature of coldest month	Winter cold	−29.3–17.4 °C
BIO14	Precipitation of driest month	Acute drought proxy	0–109 mm
BIO17	Precipitation of driest quarter	Seasonal drought proxy	0–340 mm

2.2 Preregistered gene panel and variant extraction

A preregistered panel of 55 genes was curated from the abiotic-stress and flowering-time literature and stratified into three functional categories: 23 drought/ABA-signaling genes (e.g. *NCED3*, *ABF2*, *ABF3*, *ABF4*, *DREB2A*, *DREB2B*, the *CBF1–3* tandem, *PYR1*, *OST1*, *RD29A/B*), 15 heat-response genes (heat-shock factors *HSFA1a/b/d*, *HSFA2*, *HSFA3*, *HSFA6b*, *HSFA7a*, *HSFB1*, and chaperones *HSP101*, *HSP90-1/2*, *HSP70-1*, small HSPs), and 17 phenology genes (*FLC*, *FRI*, *FT*, *SOC1*, *CO*, *GI*, *FLM*, *SVP*, *PHYB*, *VIN3*, and others). TAIR10 coordinates

were resolved by batch lookup against Ensembl Plants (Araport11 annotation) and cross-checked against the curated symbols; the 12 mismatches were all recognized synonyms (e.g. DREB1A=CBF3, LTI78=RD29A, MAF1=FLM). Each gene was extended by ±10 kb flanks to capture proximal cis-regulatory variation.

Genotypes were extracted from the GMI–MPI v3.1 imputed binary SNP matrix (10.7 M SNPs × 1,135 accessions). Because *A. thaliana* accessions are near-fully homozygous inbred lines, genotypes are effectively haploid and were coded as ALT-allele dosage (0 = homozygous reference, 1 = homozygous alternate; no heterozygous class). Within the 55 gene windows, 89,760 candidate SNPs were reduced to **15,686** after quality control (biallelic; missingness ≤ 10%; minor-allele frequency ≥ 0.05), with 74,074 dropped for low MAF and none for missingness (the matrix is imputed). Every gene retained ≥ 83 SNPs. Post-QC MAF ranged from 0.050 (median 0.173) to 0.500.

2.3 Population-structure diagnostics

Population structure was diagnosed by principal-component analysis (PCA) on the gene-panel genotypes. To prevent tight clusters of correlated SNPs within the ±10 kb windows from dominating the axes, we first applied PLINK-style [34, 35] sliding-window linkage-disequilibrium (LD) pruning (window = 50 SNPs, step = 5, pairwise r^2 threshold = 0.2), which retained 2,134 of 15,686 SNPs (13.6%). The pruned matrix was standardized (centered and scaled to unit variance) and decomposed with a randomized SVD solver (`scikit-learn`, 10 components, random state 42). The full 15,686-SNP matrix was preserved for association mapping. Because PCA here is computed on a targeted candidate-gene panel rather than genome-wide random markers, the resulting axes are used strictly as covariates to *control* for structure, not as a definitive genome-wide ancestry estimate.

2.4 Ancestry-controlled association and meta-analysis

Genotype–climate associations were estimated for all 15,686 SNPs against each of BIO5, BIO6, BIO14, and BIO17 under a two-tier structure-control scheme motivated by the mixed-model and principal-component corrections standard in GWAS [29–31, 36, 37]. First, analyses were stratified *within* each of the nine major ancestry groups with $N \geq 50$ accessions (Central Europe 181, Germany 171, South Sweden 156, Admixed 136, Western Europe 117, Spain 110, Italy/Balkan/Caucasus 92, Asia 79, North Sweden 64); the Relict group ($n = 25$) was excluded from association testing for low power but retained for selection. Second, within each group the per-SNP model

$$Y = \beta_0 + \beta_{\mathrm{SNP}}\,\mathrm{SNP} + \sum_{k=1}^{5} \gamma_k\,\mathrm{PC}_k + \varepsilon \tag{1}$$

was fit for each climate variable Y. Using the Frisch–Waugh–Lovell theorem, Y and every SNP were residualized on $Z = [\mathbf{1}, \mathrm{PC}_1, \ldots, \mathrm{PC}_5]$, and β_{SNP} was recovered as the univariate slope of residualized Y on residualized SNP—identical to full OLS but vectorized across all SNPs (residual df $= N_{\mathrm{group}} - 7$). Monomorphic SNPs within a group were set to missing and dropped from that group's contribution.

Group-level estimates were combined by fixed-effect inverse-variance-weighted (IVW) meta-analysis [38]: with weights $w_g = 1/\mathrm{SE}_g^2$,

$$\hat{\beta}_{\mathrm{meta}} = \frac{\sum_g w_g \hat{\beta}_g}{\sum_g w_g}, \qquad \mathrm{SE}_{\mathrm{meta}} = \sqrt{\frac{1}{\sum_g w_g}}, \qquad z = \frac{\hat{\beta}_{\mathrm{meta}}}{\mathrm{SE}_{\mathrm{meta}}}, \tag{2}$$

with two-sided p-values from the standard normal. Between-group heterogeneity was quantified by Cochran's Q and $I^2 = \max\left(0, \frac{Q-\mathrm{df}}{Q}\right) \times 100$ [39]. Multiple testing was controlled per climate

variable with the Benjamini–Hochberg false-discovery rate (FDR); Bonferroni counts are also reported. Residual population-structure confounding was assessed with the genomic inflation factor λ_{GC} [40], contrasting the PC-controlled within-group meta-analysis against a naive pooled model with no PC adjustment.

2.5 Maximin diversity selection

Accession selection maximized diversity across three spaces simultaneously rather than ranking by association strength. For the 1,131 candidates with complete climate data, three Euclidean distance matrices were computed: (i) *genetic* distance on the per-SNP standardized 15,686-SNP matrix; (ii) *climate* distance on standardized BIO5/BIO6/BIO14/BIO17; and (iii) *ancestry* distance on the top three PCs. Each matrix was normalized by its maximum to $[0, 1]$ and combined with equal weights, $D_{\text{joint}} = (D_{\text{gen}} + D_{\text{clim}} + D_{\text{anc}})/3$. A deterministic greedy maximin algorithm was then run: initialize with the pair of accessions at maximum joint distance, and iteratively add the accession that maximizes its minimum joint distance to the already-selected set until 24 are chosen (ties broken by lowest index). This "max–min dispersion" criterion spreads the panel to the periphery and interior of the joint space, ensuring coverage of divergent lineages and climatic extremes.

2.6 Factorial drought × heat design

The 24 nominated accessions were assigned to a 2×2 factorial experiment crossing **watering** (well-watered, 60% soil water content [SWC]; vs. drought, 30% SWC) with **temperature** (control, 22 °C day / 18 °C night; vs. heat, 38 °C for 6 h/day during the light period, otherwise 22 °C), yielding four treatments: Control (C), Drought (D), Heat (H), and Combined (DH). The design is a randomized complete block design (RCBD) with four blocks (benches); each block contains exactly one replicate of every accession × treatment combination (96 plants arranged in an 8×12 grid), for $24 \times 4 \times 4 = 384$ plants and 16 replicates per accession. Within each block, the 96 accession × treatment cells were independently assigned to grid positions with `numpy.random.default_rng(42).permutation`, so that physical positions are randomized independently per block to absorb micro-environmental gradients. The design was validated programmatically (see Section 3.6). All steps used a fixed random seed (42) for full reproducibility.

3 Results

3.1 Climatic breadth of the source panel

Georeferencing the 1,135 accessions against WorldClim confirmed the wide environmental envelope of the species (Figure 1). Across the georeferenced pool, the heat proxy BIO5 ranged from 7.9 to 36.0 °C, winter cold BIO6 from −29.3 to 17.4 °C, and the drought proxies from 0 to 109 mm (BIO14) and 0 to 340 mm (BIO17). This range spans hyper-arid, near-zero dry-season precipitation sites (e.g. the Cape Verde Relict Cvi-0) through cold continental Asian and Scandinavian provenances, providing the raw variation that the selection step exploits.

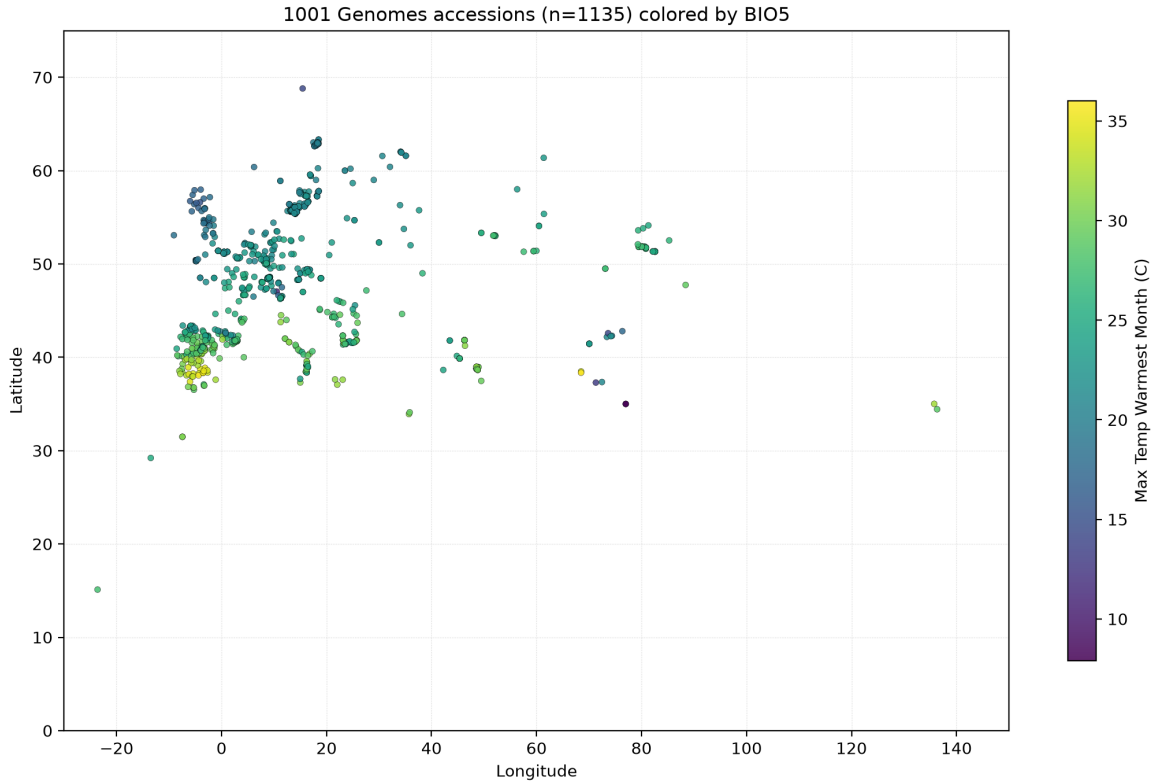
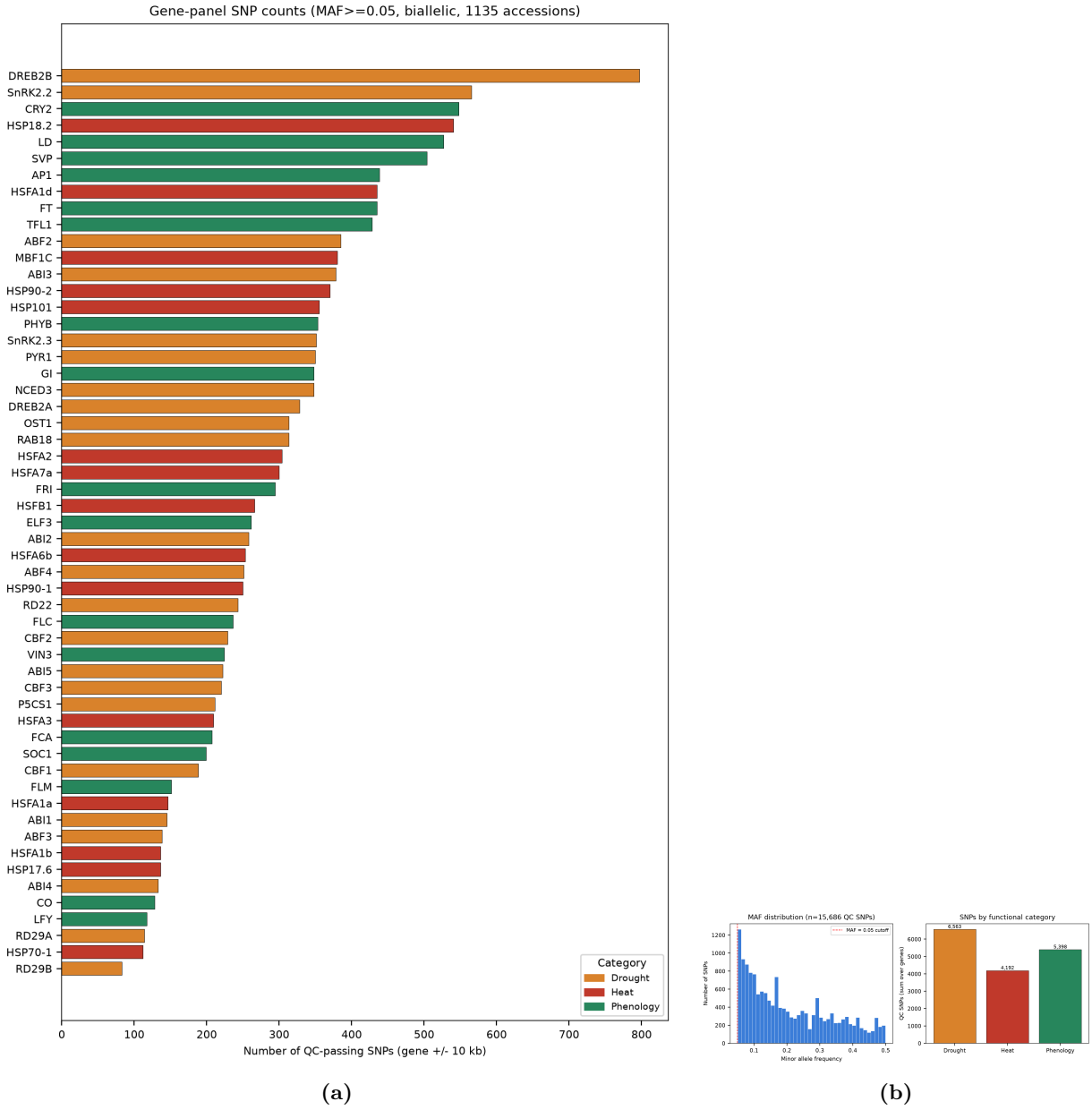


Figure 1: Geographic distribution of the 1,135 1001 Genomes accessions, colored by the heat proxy BIO5 (maximum temperature of the warmest month). The panel spans the Eurasian and North African native range plus introduced North American sites, capturing a broad temperature and aridity gradient that underpins climate-informed selection.

3.2 Gene panel and variant catalog

The preregistered 55-gene panel resolved cleanly to the five nuclear chromosomes and yielded a dense, evenly distributed variant catalog: after QC, 15,686 SNPs remained, with every gene contributing $\geq$ 83 variants (Figure 2a). SNP counts per gene reflected window size and local polymorphism, ranging from 83 (*RD29B*) to 797 (*DREB2B*); drought, heat, and phenology categories were all well represented. The post-QC minor-allele-frequency spectrum was typical of a common-variant panel, with a median MAF of 0.173 and no missing calls owing to imputation (Figure 2b). This catalog provides both the association substrate and the standardized genotype space used for maximin selection.



(a) **(b)**

Figure 2: Variant catalog for the 55-gene panel. (a) Quality-controlled SNP counts per gene, colored by functional category (drought, heat, phenology). (b) Minor-allele-frequency distribution of the 15,686 retained SNPs (median 0.173).

3.3 Population-structure diagnostics

LD pruning retained 2,134 informative SNPs, on which PCA recovered the expected continental structure of the panel. The leading axes explained modest but interpretable variance—PC1 6.68%, PC2 4.27%, PC3 2.89%, PC4 2.46%, with the top ten components summing to 24.6% (Table 2; Figure 3)—and the scree plot showed an elbow after roughly PC4–PC5, justifying the use of five PCs as association covariates. In the PC1–PC2 plane the Asian, Spanish, and Swedish groups separated along recognizable geographic gradients, while all 25 Relict accessions sat distinctly apart from the main European/admixed cloud (Figure 4), consistent with their status as a deeply divergent, refugial lineage [17, 18]. The PC ordering matched the accession list exactly, and the diagnostics confirm that structure is strong enough to require correction in the association step.

Table 2: Population-structure summary. Left: composition of the panel by ADMIXTURE genetic group. Right: variance explained by the leading principal components of the LD-pruned marker set.

Genetic group	n **accessions**
Central Europe	184
Germany	171
South Sweden	156
Admixed	137
Western Europe	117
Spain	110
Italy, Balkan and Caucasus	92
Asia	79
North Sweden	64
Relict	25
Total	**1,135**

PC	**Var. (%)**	**Cum. (%)**
PC1	6.68	6.68
PC2	4.27	10.95
PC3	2.89	13.84
PC4	2.46	16.30
PC5	1.87	18.17
PC6	1.62	19.79
PC7	1.45	21.24
PC8	1.33	22.57
PC9	1.05	23.62
PC10	0.99	24.62

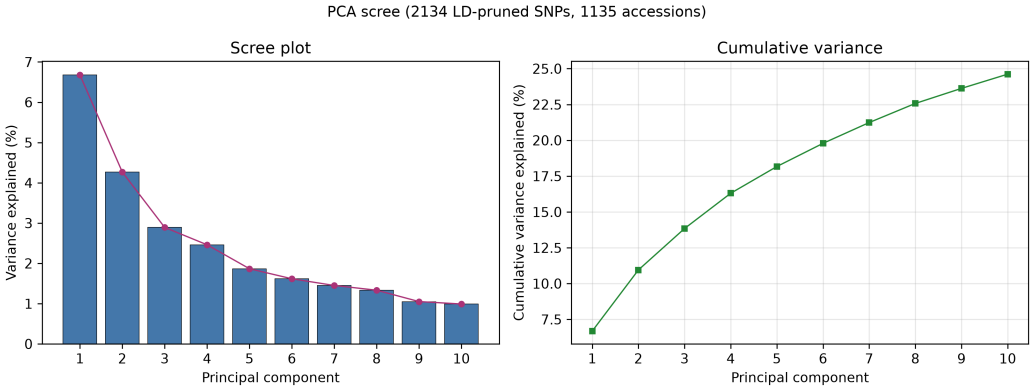


Figure 3: Scree plot of per-PC and cumulative variance explained by PCA on the LD-pruned (2,134-SNP) genotype matrix. The elbow after PC4–PC5 supports using five PCs as covariates in the association model.

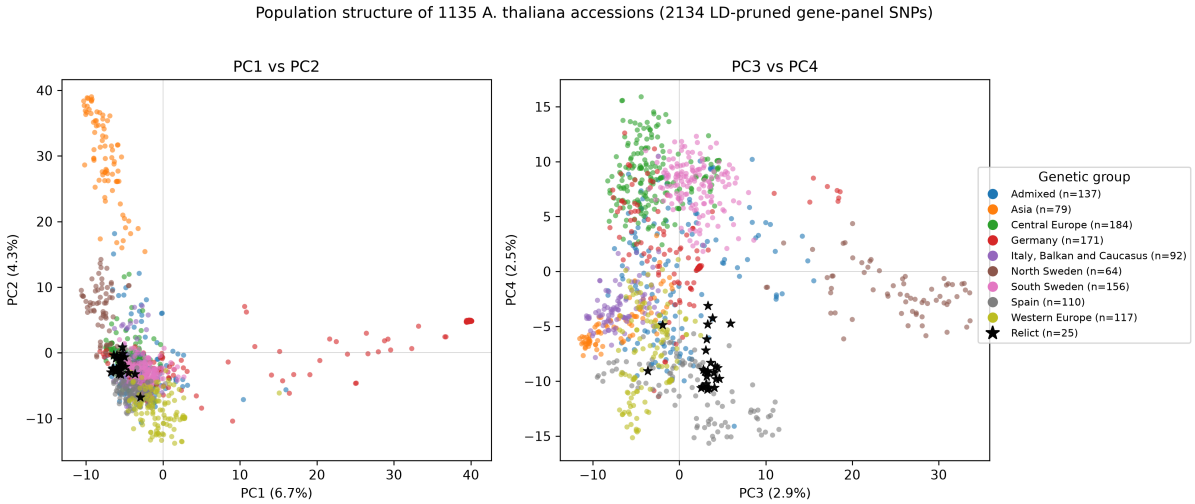


Figure 4: Principal-component structure of the 1,135 accessions, colored by ADMIXTURE genetic group, for PC1–PC2 (left) and PC3–PC4 (right). The 25 Relict accessions (highlighted) separate from the main European/admixed cloud, and continental groups (Asia, Spain, Sweden) resolve along the leading axes.

3.4 Ancestry-controlled associations and locus-level hypotheses

Association testing across nine groups and four climate variables produced signals concentrated at biologically coherent loci. The two-tier structure control was effective: the genomic inflation factor of the PC-controlled within-group meta-analysis was 2.53 (BIO5), 2.67 (BIO6), 2.58 (BIO14), and 2.75 (BIO17), a 6–15× reduction from the naive pooled no-PC model (20.9, 40.1, 16.0, 18.6, respectively). We emphasize that a residual $\lambda_{GC} > 1$ is *expected and not a red flag* here, because all 15,686 SNPs lie inside 55 candidate climate-adaptation genes in strong local LD—violating the genome-wide null of "most markers unassociated" on which λ_{GC} is premised—so the residual elevation reflects genuine polygenic enrichment at candidate loci rather than uncontrolled stratification (Figure 6). At FDR < 0.05, 901 (BIO5), 995 (BIO6), 1,131 (BIO14), and 1,146 (BIO17) SNPs were significant; the corresponding Bonferroni counts were 178, 19, 116, and 96.

The Manhattan profiles (Figure 5) and the ranked hits (Table 3) point to a small set of high-confidence, falsifiable locus-level hypotheses. For the heat proxy BIO5, the strongest signals fell in *HSFA2*—the heat-shock factor required for acquired thermotolerance [12]—and in the florigen *FT*, the canonical thermosensory flowering gene; the cold/dehydration *CBF1–3* tandem also surfaced. For winter cold (BIO6), the vernalization repressor *FLC* [16] and the ABA regulator *ABF2* led, alongside *HSFB1* and the autonomous-pathway gene *FCA*. For both drought proxies (BIO14, BIO17), the ABA-biosynthesis enzyme *NCED3* [7] and the seed/ABA regulator *ABI3* produced the strongest drought-category hits, with the thermotolerance chaperone *HSP101* [13] and the thermosensory flowering repressor *FLM* also prominent—consistent with the tight coupling between water availability, heat, and flowering time in Mediterranean climates. Each hypothesis is reported with its effect size, standard error, p-value, FDR, and I^2 heterogeneity (Table 3); the generally high I^2 at the top loci indicates that effect magnitudes—though directionally consistent—vary among ancestry groups, itself an argument for a diverse experimental panel able to test G×E.

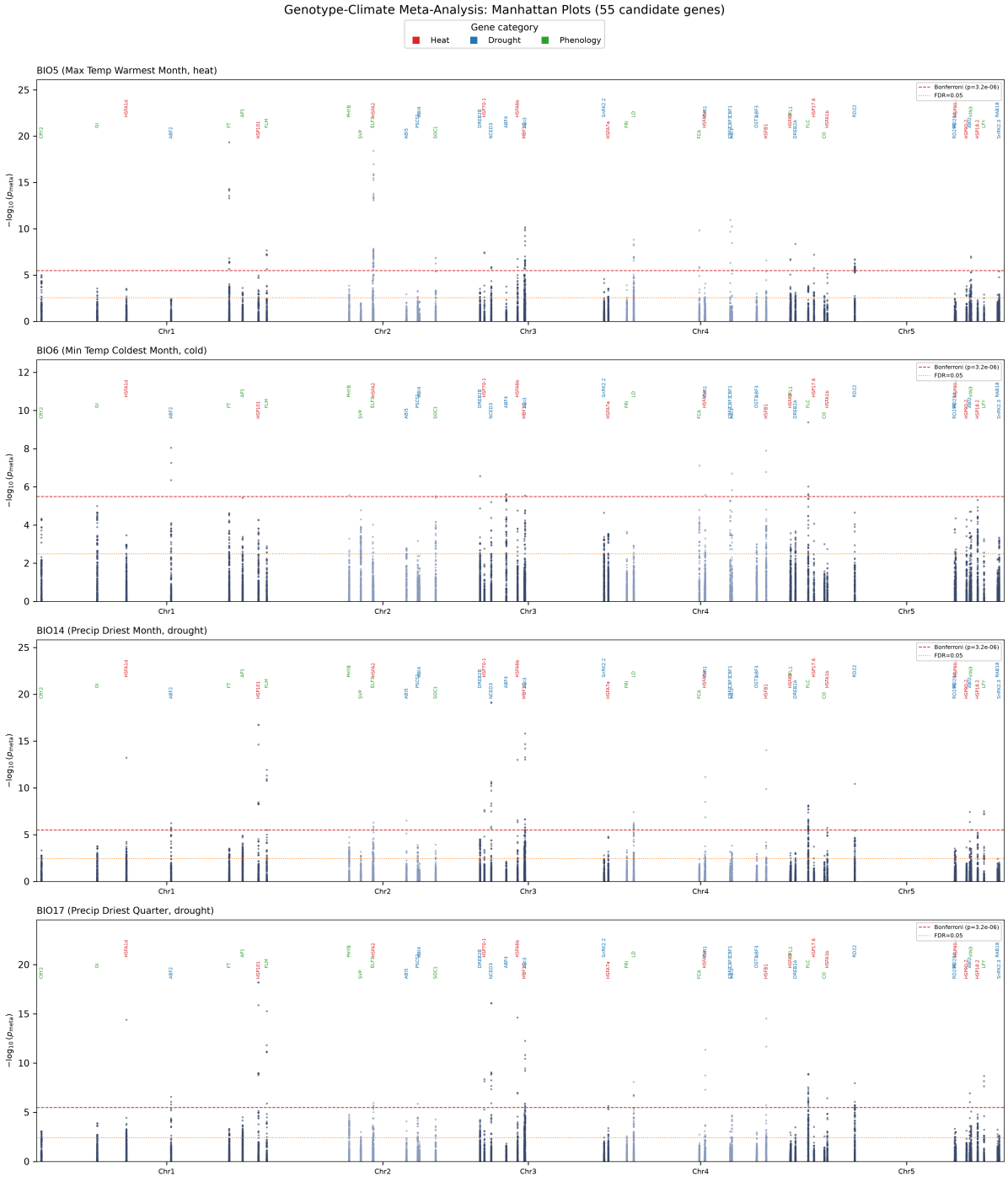


Figure 5: Manhattan plots of the ancestry-controlled IVW meta-analysis for the four bioclimatic variables. Points are SNPs positioned by genomic coordinate within the 55 gene windows and colored by functional category; horizontal lines mark the Benjamini–Hochberg FDR and Bonferroni thresholds. Peak genes are labeled.

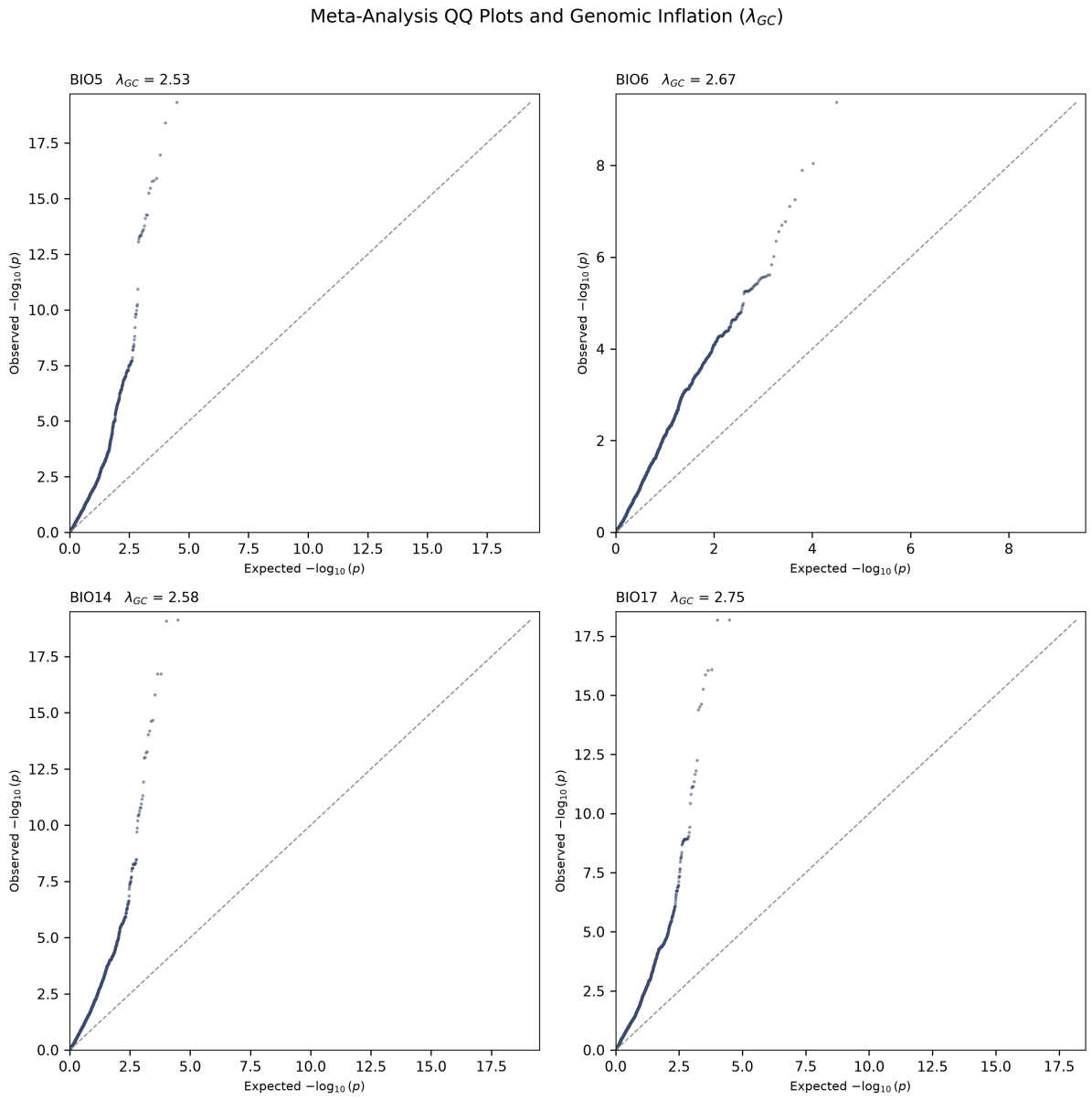


Figure 6: Quantile–quantile plots for the four climate variables, annotated with the genomic inflation factor λ_{GC}. The null-aligned bulk with a strong true-signal tail is consistent with a candidate-gene panel in local LD (see text).

Table 3: Representative locus-level hypotheses with quantified uncertainty. For each bioclimatic variable, top SNPs (by meta-analysis p) are shown with the annotated gene and category, meta-analytic effect ($\hat{\beta}_{\text{meta}}$, in climate units per ALT allele) and standard error, two-sided p, Benjamini–Hochberg FDR, Cochran's I^2 (% between-group heterogeneity), and the number of contributing groups. High I^2 flags effect-size variation among ancestry groups.

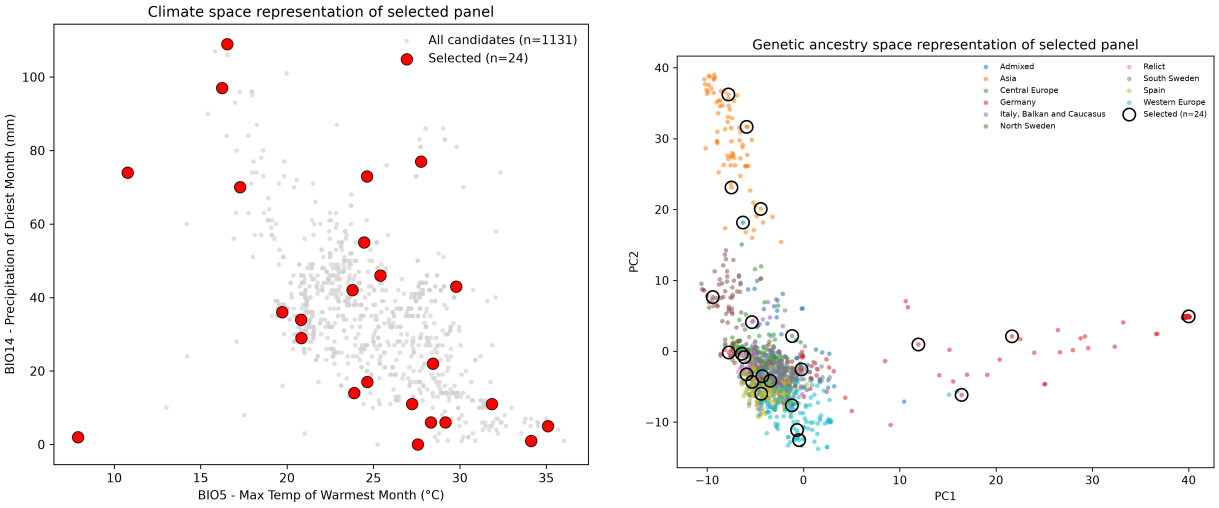
Clim.	SNP	Gene	Cat.	$\hat{\beta}$	SE	p	FDR	I^2	n_g
BIO5	Chr1:24334966	*FT*	Phen.	−0.75	0.08	4.8×10^{-20}	7.5×10^{-16}	90.9	9
BIO5	Chr2:11142985	*HSFA2*	Heat	0.78	0.09	4.0×10^{-19}	3.1×10^{-15}	90.0	9
BIO5	Chr4:13022353	*CBF1–3*	Drought	0.75	0.11	1.2×10^{-11}	8.5×10^{-9}	91.3	9
BIO5	Chr3:9001921	*ABI3*	Drought	0.48	0.07	6.9×10^{-11}	4.5×10^{-8}	0.0	9
BIO6	Chr5:3173943	*FLC*	Phen.	0.89	0.14	4.2×10^{-10}	6.7×10^{-6}	71.5	8
BIO6	Chr1:17171855	*ABF2*	Drought	−0.95	0.17	9.1×10^{-9}	6.7×10^{-5}	60.7	9
BIO6	Chr4:17441431	*HSFB1*	Heat	−1.17	0.21	1.3×10^{-8}	6.7×10^{-5}	67.5	5
BIO6	Chr4:9213695	*FCA*	Phen.	−2.26	0.42	7.8×10^{-8}	2.5×10^{-4}	31.9	7
BIO14	Chr3:4843672	*NCED3*	Drought	−5.44	0.60	7.6×10^{-20}	6.7×10^{-16}	90.6	8
BIO14	Chr1:27934049	*HSP101*	Heat	−8.63	1.02	1.9×10^{-17}	7.5×10^{-14}	71.1	8
BIO14	Chr3:9009321	*ABI3*	Drought	−4.72	0.57	1.6×10^{-16}	5.1×10^{-13}	89.8	9
BIO17	Chr1:27934046	*HSP101*	Heat	−27.20	3.06	6.6×10^{-19}	5.2×10^{-15}	67.5	8
BIO17	Chr3:4843672	*NCED3*	Drought	−15.54	1.87	8.2×10^{-17}	3.5×10^{-13}	88.4	8
BIO17	Chr1:28966182	*FLM*	Phen.	−22.45	2.77	5.6×10^{-16}	1.5×10^{-12}	86.8	9

3.5 The 24 nominated accessions

Greedy maximin optimization of the equally weighted genotype/climate/ancestry distance nominated the 24 accessions in Table 4. The panel achieves the design goal of broad, balanced diversity: it spans **all nine genetic groups**, including **two Relict** accessions (the Cape Verde island line Cvi-0 and the Iberian IP-Sne-0), and it captures the climatic extremes of the source pool—the selected minimum and maximum equal the full-pool range for BIO6, BIO14, and BIO17, and approach it for BIO5. The minimum pairwise joint distance among the 24 was 0.384 and the mean was 0.564 (on the $[0, 1]$-normalized scale), confirming that no two accessions are redundant and that the set is well dispersed. Figures 7a and 7b show the selected accessions highlighted against the candidate pool in climate space (BIO5 vs. BIO14) and ancestry space (PC1 vs. PC2), respectively, and Figure 8 shows the clustered joint-distance matrix among the 24, visually confirming near-uniform separation.

Table 4: The 24 nominated accessions for the drought-by-heat common-garden experiment, in maximin selection order. Coordinates are collection sites; climate values are WorldClim v2.1 normals at those coordinates (BIO5/BIO6 in °C, BIO14/BIO17 in mm). The panel spans all nine genetic groups, including two Relict lineages.

#	Accession	CS no.	Country	Genetic group	Lat	Lon	BIO5	BIO6	BIO14	BIO17
1	UKNW06-003	CS78792	United Kingdom	Western Europe	54.50	−3.00	16.5	−1.5	109	340
2	Kondara	CS76532	Tajikistan	Asia	38.48	68.49	34.1	−2.9	1	8
3	Omn-5	CS77146	Sweden	North Sweden	62.93	18.34	19.7	−9.8	36	122
4	Tol-3	CS79020	United States	Germany	41.66	−83.56	29.8	−8.4	43	153
5	Cvi-0	CS76789	Cabo Verde	Relict	15.11	−23.62	27.6	17.4	0	0
6	Balan-1	CS76687	Russian Fed.	Italy, Balkan and Caucasus	55.36	61.41	23.9	−17.3	14	49
7	IP-Bis-0	CS76711	Spain	Spain	42.49	0.54	24.5	−3.6	55	214
8	Fei-0	CS76412	Portugal	Western Europe	40.92	−8.54	24.6	4.7	17	88
9	CSHL-5	CS76779	United States	Germany	40.86	−73.47	27.8	−5.0	77	262
10	Kas-2	CS78905	India	Asia	35.00	77.00	7.9	−29.3	2	11
11	Radk-2	CS78928	Russian Fed.	Asia	61.59	35.11	20.8	−12.7	29	95
12	Pi-0	CS76572	Austria	Central Europe	47.04	10.51	10.8	−12.0	74	246
13	IP-San-10	CS77232	Spain	Admixed	38.33	−3.51	35.1	1.5	5	31
14	TV-4	CS78771	Sweden	Germany	55.58	14.33	20.8	−1.7	34	115
15	Bela-2	CS76697	Slovakia	Central Europe	48.47	18.94	23.8	−6.6	42	133
16	Kil-0	CS76526	United Kingdom	Germany	55.64	−5.66	17.3	3.0	70	231
17	IP-Sne-0	CS77258	Spain	Relict	37.09	−3.38	27.2	−6.5	11	56
18	DIR-9	CS76796	France	Germany	48.54	4.32	25.4	−0.2	46	144
19	Ak-1	CS76431	Germany	Admixed	48.07	7.63	24.6	−1.2	73	232
20	VED-10	CS78839	France	Admixed	43.74	3.89	28.4	2.5	22	102
21	Altai-5	CS76433	China	Asia	47.75	88.40	29.2	−21.9	6	20
22	Yo-0	CS76633	United States	Germany	37.45	−119.35	28.3	−3.4	6	24
23	UKID116	CS78780	United Kingdom	Admixed	56.73	−5.98	16.2	3.3	97	322
24	IP-Boa-0	CS76714	Spain	Spain	40.40	−3.88	31.8	1.4	11	51



(a) **(b)**

Figure 7: Maximin selection in two of the three optimization spaces. (a) Climate space (BIO5 vs. BIO14): the 24 selected accessions (red) tile the periphery and interior of the candidate cloud. (b) Ancestry space (PC1 vs. PC2): selected accessions (circled) are drawn from every genetic group, including the divergent Relict lineage.

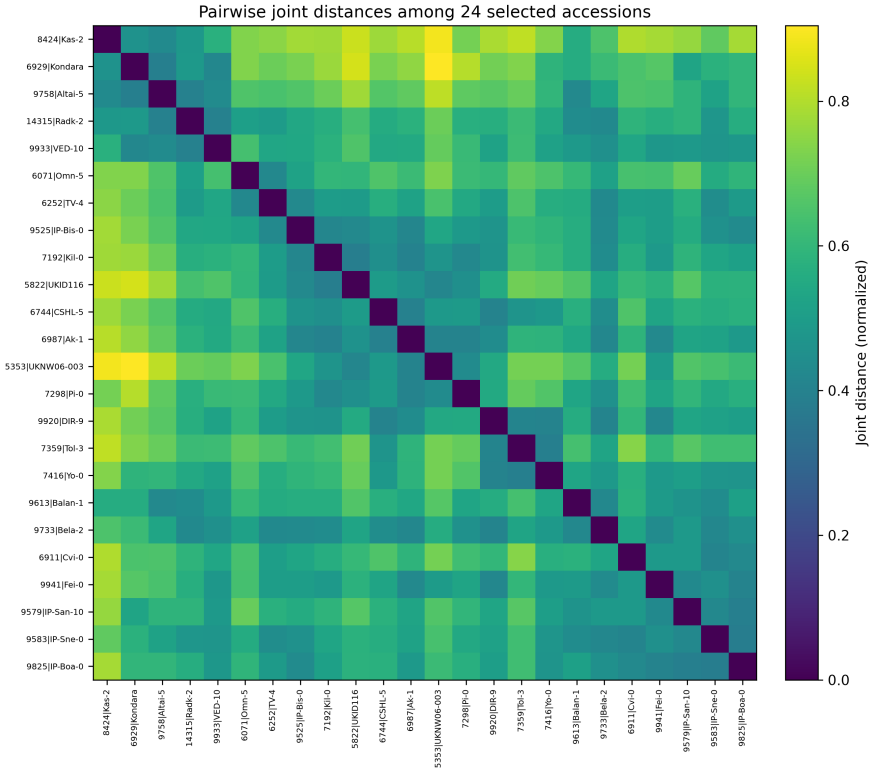


Figure 8: Clustered heatmap of pairwise joint (genotype+climate+ancestry) distances among the 24 nominated accessions. Uniformly high off-diagonal distances (minimum 0.384, mean 0.564 on the normalized scale) confirm that the panel contains no redundant pairs.

3.6 Factorial drought × heat layout

The 24 accessions were assigned to the 2 × 2 factorial RCBD summarized in Table 5. The four treatments (Control, Drought, Heat, Combined) cross well-watered vs. drought irrigation with

control vs. heat-shock temperature, so that the Combined arm and its three matched controls permit formal partitioning of drought main effects, heat main effects, and—critically—the drought × heat interaction. The design places one replicate of each of the 96 accession × treatment cells in each of four blocks (8 × 12 benches), for 384 plants, 16 per accession and 96 per treatment, with positions randomized independently within each block (Figure 9). All balance and integrity constraints passed: 384 unique plant IDs, exactly 96 plants per block and per treatment, exactly 16 plants per accession, one replicate of every accession × treatment combination per block, and zero duplicate grid positions (Table 5).

Table 5: Factorial drought × heat design (RCBD) and its validation. SWC, soil water content.

(a) 2 × 2 treatment structure

Treatment	Watering	Temperature
Control (C)	Well-watered, 60% SWC	22/18 °C
Drought (D)	Drought, 30% SWC	22/18 °C
Heat (H)	Well-watered, 60% SWC	38 °C for 6 h/day
Combined (DH)	Drought, 30% SWC	38 °C for 6 h/day

(b) Design parameters

Quantity	Value
Accessions	24
Treatments	4
Blocks	4
Reps / cell	4
Plants / block	96
Plants / treatment	96
Plants / accession	16
Total plants	**384**
Duplicate positions	0

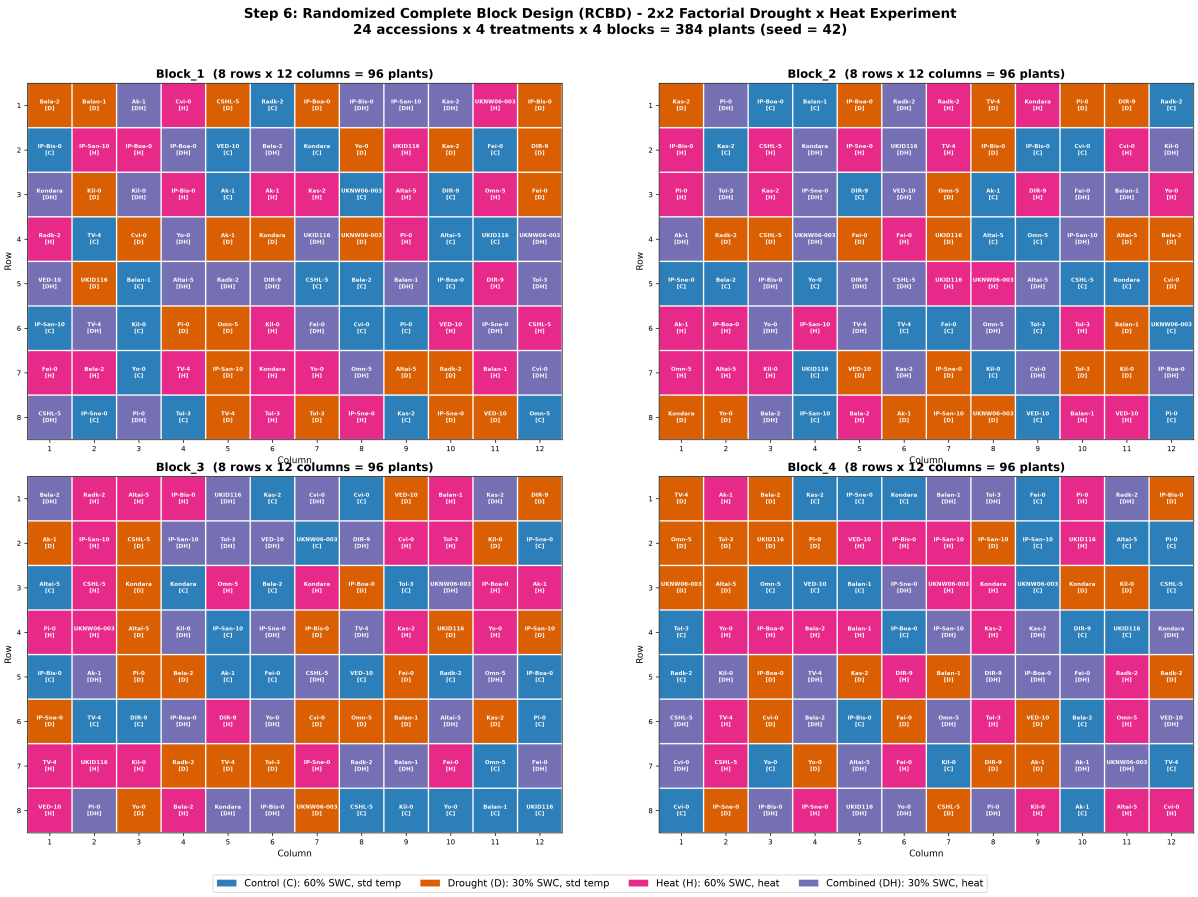


Figure 9: Randomized complete block layout for the four benches (blocks). Each of the four 8×12 grids holds one replicate of every accession × treatment combination; cells are colored by treatment (Control, Drought, Heat, Combined) and labeled by accession. Positions were randomized independently per block (seed 42) to control micro-environmental gradients.

4 Discussion

4.1 Statistical power and the analysis plan

The delivered design is powered for its primary questions. Each of the four treatments is represented by 96 plants (24 accessions × 4 blocks), which affords high power to detect watering and temperature main effects and their interaction across the panel; the drought × heat interaction contrast—the biologically novel quantity [3, 4]—is estimated from all 384 plants. The planned primary analysis is a linear mixed-effects model, `response` ∼ `watering * temperature * genetic_group + (1|block)`, in which the random block effect absorbs bench-level micro-environmental variation and shrinks noisy per-accession contrasts; a two- or three-way ANOVA with block as a fixed factor is the simpler alternative when variance components are not of interest. Effects will be reported as estimates with 95% confidence intervals alongside p-values, and accession- or trait-level multiplicity will be controlled by Benjamini–Hochberg FDR. Per-accession G×E contrasts rely on four biological replicates per cell (one per block); these are best interpreted through the mixed-model shrinkage rather than as independent per-accession tests, and the panel's deliberate ancestry and climate spread maximizes the leverage available for detecting genotype-dependent stress responses [32, 33].

4.2 Biological interpretation of the locus-level hypotheses

The association results are notable less for discovery than for their face validity, which is exactly what a candidate-panel design should provide: strong, ancestry-controlled signals recovered the textbook regulators of each stress axis. The appearance of *HSFA2* and *HSP101* as leading hits for the heat and aridity proxies is consistent with their established, non-redundant roles in acquired thermotolerance and survival of severe heat [12–14], while *NCED3* and *ABI3* for the drought proxies reflect the centrality of ABA biosynthesis and signaling to water-deficit adaptation [7–9]. The recurrence of phenology genes—*FT* and *FLM* with temperature, *FLC* with winter cold—echoes the well-documented coupling of flowering time to climate in this species, whereby drought escape via accelerated flowering is itself an adaptive strategy [16, 25, 41]. The generally high between-group heterogeneity (I^2) at the top loci is itself informative: allelic effects on climate association differ among ancestry groups, implying that a single-population experiment would give a partial and potentially misleading picture, and reinforcing the rationale for a maximally diverse common-garden panel. Because these are provenance-climate associations rather than experimental phenotypes, they constitute *hypotheses*—for example, that ALT alleles at the *NCED3* and *HSP101* loci confer measurable differences in water-use or thermotolerance phenotypes—that the factorial experiment is expressly designed to test.

4.3 Why maximin rather than top-hit selection

Selecting accessions by association strength alone would have concentrated the panel on a few haplotypes at the extreme of each climate gradient and under-sampled the ancestry axis, undermining both generalizability and the ability to separate causal alleles from their genomic background [29, 32]. The maximin criterion instead treats diversity as the objective, and the outcome bears this out: the 24 accessions cover all nine genetic groups and both Relict sublineages, and they reproduce the full pool range on three of four climate axes. This breadth is the strongest guarantor that treatment effects estimated in the common garden will generalize across the species and that G×E, if present, will be detectable. The equal 1:1:1 weighting of the genotype, climate, and ancestry distances is a transparent default; investigators who wish to emphasize, say, climatic contrast could re-weight the components without altering the rest of the pipeline.

4.4 Limitations

Several caveats bound the interpretation. First, provenance climate is a coarse, static proxy: WorldClim normals describe 1970–2000 averages at the collection coordinate and cannot capture microhabitat, phenological timing of stress exposure, or recent change, so genotype–climate associations index long-run adaptation rather than the imposed experimental stresses [22]. Second, the association analysis is deliberately restricted to a candidate panel in strong local LD; this maximizes prior biological relevance but means the reported SNPs tag gene regions rather than pinpoint causal variants, and the elevated λ_{GC} must be read in that light rather than as a genome-wide confounding metric [40]. Third, the fixed-effect IVW meta-analysis assumes independent group-level estimates and summarizes heterogeneity post hoc via I^2; a random-effects model would widen intervals where I^2 is high. Fourth, the Relict lineage was excluded from association testing for low power ($n = 25$) even though two Relicts were retained for the experiment—so no climate-association hypotheses are offered specifically for those accessions. Fifth, ancestry-group labels derive from ADMIXTURE assignments with a 60% ancestry threshold and are therefore coarse; the "Admixed" class in particular is heterogeneous [17]. Finally, four biological replicates per accession × treatment cell provide robust power for panel-level effects but limited power for individual per-accession contrasts, which should be treated as exploratory and validated in follow-up.

4.5 Conclusion

We have delivered a complete, reproducible pipeline—from the 1001 Genomes and WorldClim resources to a validated bench layout—that nominates 24 *A. thaliana* accessions for a factorial drought-by-heat common-garden experiment. By controlling genotype–climate associations for ancestry within groups and meta-analyzing across them, the pipeline generates well-calibrated locus-level hypotheses centered on the canonical drought (*NCED3*, *ABI3*, *ABF2*), heat (*HSFA2*, *HSP101*), and phenology (*FT*, *FLC*, *FLM*) machinery. By selecting on maximin diversity rather than association rank, it produces a panel that spans the species' ancestry and climatic breadth. The resulting 2×2 factorial RCBD, with matched controls and 16 replicates per accession, is ready to execute and analyze under the specified mixed-effects framework. Together these products convert genomic and climatic data into an actionable, hypothesis-driven experiment on the genetics of combined-stress tolerance.

Data and code availability

All analyses used publicly available data: the 1001 Genomes GMI–MPI v3.1 imputed SNP matrix and accession metadata [17], and WorldClim v2.1 1970–2000 bioclimatic normals [22]. The workflow is organized as numbered, seed-fixed (42) scripts producing the accession–climate table, the gene-panel genotype matrix and per-SNP metadata, the LD-pruned matrix and principal components, the full meta-analysis table and top-loci summary, the 24-accession nomination, and the 384-plant experimental layout, each accompanied by a machine-readable JSON summary and diagnostic figures reproduced in this report.

References

[1] Ron Mittler. Abiotic stress, the field environment and stress combination. *Trends in Plant Science*, 11(1):15–19, January 2006. ISSN 1360-1385. doi: 10.1016/j.tplants.2005.11.002. URL http://dx.doi.org/10.1016/j.tplants.2005.11.002.

[2] Sara I. Zandalinas, Felix B. Fritschi, and Ron Mittler. Global warming, climate change, and environmental pollution: Recipe for a multifactorial stress combination disaster. *Trends in Plant Science*, 26(6):588–599, 2021. ISSN 1360-1385. doi: 10.1016/j.tplants.2021.02.011. URL http://dx.doi.org/10.1016/j.tplants.2021.02.011.

[3] Ludmila Rizhsky, Hongjian Liang, Joel Shuman, Vladimir Shulaev, Sholpan Davletova, and Ron Mittler. When defense pathways collide. the response of arabidopsis to a combination of drought and heat stress . *Plant Physiology*, 134(4):1683–1696, April 2004. ISSN 0032-0889. doi: 10.1104/pp.103.033431. URL http://dx.doi.org/10.1104/pp.103.033431.

[4] Nobuhiro Suzuki, Rosa M. Rivero, Vladimir Shulaev, Eduardo Blumwald, and Ron Mittler. Abiotic and biotic stress combinations. *New Phytologist*, 203(1):32–43, April 2014. ISSN 1469-8137. doi: 10.1111/nph.12797. URL http://dx.doi.org/10.1111/nph.12797.

[5] Outi Savolainen, Martin Lascoux, and Juha Merilä. Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11):807–820, October 2013. ISSN 1471-0064. doi: 10.1038/nrg3522. URL http://dx.doi.org/10.1038/nrg3522.

[6] Tadeusz J. Kawecki and Dieter Ebert. Conceptual issues in local adaptation. *Ecology Letters*, 7(12):1225–1241, November 2004. ISSN 1461-0248. doi: 10.1111/j.1461-0248.2004.00684.x. URL http://dx.doi.org/10.1111/j.1461-0248.2004.00684.x.

[7] Satoshi Iuchi, Masatomo Kobayashi, Teruaki Taji, Masaaki Naramoto, Motoaki Seki, Tomohiko Kato, Satoshi Tabata, Yoshitaka Kakubari, Kazuko Yamaguchi-Shinozaki, and Kazuo Shinozaki. Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in arabidopsis. *The Plant Journal*, 27(4):325–333, August 2001. ISSN 1365-313X. doi: 10.1046/j.1365-313x.2001.01096.x. URL http://dx.doi.org/10.1046/j.1365-313x.2001.01096.x.

[8] Kazuko Yamaguchi-Shinozaki and Kazuo Shinozaki. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annual Review of Plant Biology*, 57(1):781–803, 2006. ISSN 1545-2123. doi: 10.1146/annurev.arplant.57.032905.105444. URL http://dx.doi.org/10.1146/annurev.arplant.57.032905.105444.

[9] Yasunari Fujita, Miki Fujita, Kazuo Shinozaki, and Kazuko Yamaguchi-Shinozaki. Aba-mediated transcriptional regulation in response to osmotic stress in plants. *Journal of Plant Research*, 124(4):509–525, March 2011. ISSN 1618-0860. doi: 10.1007/s10265-011-0412-3. URL http://dx.doi.org/10.1007/s10265-011-0412-3.

[10] Yoh Sakuma, Kyonoshin Maruyama, Yuriko Osakabe, Feng Qin, Motoaki Seki, Kazuo Shinozaki, and Kazuko Yamaguchi-Shinozaki. Functional analysis of anarabidopsistranscription factor, dreb2a, involved in drought-responsive gene expression. *The Plant Cell*, 18(5):1292–1309, April 2006. ISSN 1532-298X. doi: 10.1105/tpc.105.035881. URL http://dx.doi.org/10.1105/tpc.105.035881.

[11] Jian-Kang Zhu. Abiotic stress signaling and responses in plants. *Cell*, 167(2):313–324, October 2016. ISSN 0092-8674. doi: 10.1016/j.cell.2016.08.029. URL http://dx.doi.org/10.1016/j.cell.2016.08.029.

[12] Yee-yung Charng, Hsiang-chin Liu, Nai-yu Liu, Wen-tzu Chi, Chun-neng Wang, Shih-hsun Chang, and Tsu-tsuen Wang. A heat-inducible transcription factor, hsfa2, is required for extension of acquired thermotolerance in arabidopsis. *Plant Physiology*, 143(1):251–262, November 2006. ISSN 1532-2548. doi: 10.1104/pp.106.091322. URL http://dx.doi.org/10.1104/pp.106.091322.

[13] Christine Queitsch, Suk-Whan Hong, Elizabeth Vierling, and Susan Lindquist. Heat shock protein 101 plays a crucial role in thermotolerance in arabidopsis. *The Plant Cell*, 12(4): 479–492, April 2000. ISSN 1532-298X. doi: 10.1105/tpc.12.4.479. URL http://dx.doi.org/10.1105/tpc.12.4.479.

[14] Naohiko Ohama, Hikaru Sato, Kazuo Shinozaki, and Kazuko Yamaguchi-Shinozaki. Transcriptional regulatory network of plant heat stress response. *Trends in Plant Science*, 22(1):53–65, January 2017. ISSN 1360-1385. doi: 10.1016/j.tplants.2016.08.015. URL http://dx.doi.org/10.1016/j.tplants.2016.08.015.

[15] Klaus-Dieter Scharf, Thomas Berberich, Ingo Ebersberger, and Lutz Nover. The plant heat stress transcription factor (hsf) family: Structure, function and evolution. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(2):104–119, February 2012. ISSN 1874-9399. doi: 10.1016/j.bbagrm.2011.10.002. URL http://dx.doi.org/10.1016/j.bbagrm.2011.10.002.

[16] Urban Johanson, Joanne West, Clare Lister, Scott Michaels, Richard Amasino, and Caroline Dean. Molecular analysis of frigida , a major determinant of natural variation in arabidopsis flowering time. *Science*, 290(5490):344–347, October 2000. ISSN 1095-9203. doi: 10.1126/science.290.5490.344. URL http://dx.doi.org/10.1126/science.290.5490.344.

[17] Carlos Alonso-Blanco, Jorge Andrade, Claude Becker, Felix Bemm, Joy Bergelson, Karsten M. Borgwardt, Jun Cao, Eunyoung Chae, Todd M. Dezwaan, Wei Ding, Joseph R. Ecker, Moises Exposito-Alonso, Ashley Farlow, Joffrey Fitz, Xiangchao Gan, Dominik G. Grimm, Angela M. Hancock, Stefan R. Henz, Svante Holm, Matthew Horton, Mike Jarsulic, Randall A. Kerstetter, Arthur Korte, Pamela Korte, Christa Lanz, Cheng-Ruei Lee, Dazhe Meng, Todd P. Michael, Richard Mott, Ni Wayan Muliyati, Thomas Nägele, Matthias Nagler, Viktoria Nizhynska, Magnus Nordborg, Polina Yu. Novikova, F. Xavier Picó, Alexander Platzer, Fernando A. Rabanal, Alex Rodriguez, Beth A. Rowan, Patrice A. Salomé, Karl J. Schmid, Robert J. Schmitz, Ümit Seren, Felice Gianluca Sperone, Mitchell Sudkamp, Hannes Svardal, Matt M. Tanzer, Donald Todd, Samuel L. Volchenboum, Congmao Wang, George Wang, Xi Wang, Wolfram Weckwerth, Detlef Weigel, and Xuefeng Zhou. 1,135 genomes reveal the global pattern of polymorphism in arabidopsis thaliana. *Cell*, 166(2):481–491, 2016. ISSN 0092-8674. doi: 10.1016/j.cell.2016.05.063. URL http://dx.doi.org/10.1016/j.cell.2016.05.063.

[18] Arun Durvasula, Andrea Fulgione, Rafal M. Gutaker, Selen Irez Alacakaptan, Pádraic J. Flood, Célia Neto, Takashi Tsuchimatsu, Hernán A. Burbano, F. Xavier Picó, Carlos Alonso-Blanco, and Angela M. Hancock. African genomes illuminate the early history and transition to selfing in arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, 114(20):5213–5218, May 2017. ISSN 1091-6490. doi: 10.1073/pnas.1616736114. URL http://dx.doi.org/10.1073/pnas.1616736114.

[19] Andrea Fulgione and Angela M. Hancock. Archaic lineages broaden our view on the history of arabidopsis thaliana. *New Phytologist*, 219(4):1194–1198, 2018. ISSN 1469-8137. doi: 10.1111/nph.15244. URL http://dx.doi.org/10.1111/nph.15244.

[20] Taiji Kawakatsu, Shao-shan Carol Huang, Florian Jupe, Eriko Sasaki, Robert J. Schmitz, Mark A. Urich, Rosa Castanon, Joseph R. Nery, Cesar Barragan, Yupeng He, Huaming Chen, Manu Dubin, Cheng-Ruei Lee, Congmao Wang, Felix Bemm, Claude Becker, Ryan O'Neil, Ronan C. O'Malley, Danjuma X. Quarless, Nicholas J. Schork, Detlef Weigel, Magnus Nordborg, Joseph R. Ecker, Carlos Alonso-Blanco, Jorge Andrade, Claude Becker, Felix Bemm, Joy Bergelson, Karsten Borgwardt, Eunyoung Chae, Todd Dezwaan, Wei Ding, Joseph R. Ecker, Moisés Expósito-Alonso, Ashley Farlow, Joffrey Fitz, Xiangchao Gan, Dominik G. Grimm, Angela Hancock, Stefan R. Henz, Svante Holm, Matthew Horton, Mike Jarsulic, Randall A. Kerstetter, Arthur Korte, Pamela Korte, Christa Lanz, Chen-Ruei Lee, Dazhe Meng, Todd P. Michael, Richard Mott, Ni Wayan Muliyati, Thomas Nägele, Matthias Nagler, Viktoria Nizhynska, Magnus Nordborg, Polina Novikova, F. Xavier Picó, Alexander Platzer, Fernando A. Rabanal, Alex Rodriguez, Beth A. Rowan, Patrice A. Salomé, Karl Schmid, Robert J. Schmitz, Ümit Seren, Felice Gianluca Sperone, Mitchell Sudkamp, Hannes Svardal, Matt M. Tanzer, Donald Todd, Samuel L. Volchenboum, Congmao Wang, George Wang, Xi Wang, Wolfram Weckwerth, Detlef Weigel, and Xuefeng Zhou. Epigenomic diversity in a global collection of arabidopsis thaliana accessions. *Cell*, 166(2):492–505, 2016. ISSN 0092-8674. doi: 10.1016/j.cell.2016.06.044. URL http://dx.doi.org/10.1016/j.cell.2016.06.044.

[21] Robert J. Hijmans, Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15):1965–1978, 2005. ISSN 1097-0088. doi: 10.1002/joc.1276. URL http://dx.doi.org/10.1002/joc.1276.

[22] Stephen E. Fick and Robert J. Hijmans. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315, May 2017. ISSN 1097-0088. doi: 10.1002/joc.5086. URL http://dx.doi.org/10.1002/joc.5086.

[23] Susanna Atwell, Yu S. Huang, Bjarni J. Vilhjálmsson, Glenda Willems, Matthew Horton, Yan Li, Dazhe Meng, Alexander Platt, Aaron M. Tarone, Tina T. Hu, Rong Jiang, N. Wayan Muliyati, Xu Zhang, Muhammad Ali Amer, Ivan Baxter, Benjamin Brachi, Joanne Chory, Caroline Dean, Marilyne Debieu, Juliette de Meaux, Joseph R. Ecker, Nathalie Faure, Joel M. Kniskern, Jonathan D. G. Jones, Todd Michael, Adnane Nemri, Fabrice Roux, David E. Salt, Chunlao Tang, Marco Todesco, M. Brian Traw, Detlef Weigel, Paul Marjoram, Justin O. Borevitz, Joy Bergelson, and Magnus Nordborg. Genome-wide association study of 107 phenotypes in arabidopsis thaliana inbred lines. *Nature*, 465(7298):627–631, March 2010. ISSN 1476-4687. doi: 10.1038/nature08800. URL http://dx.doi.org/10.1038/nature08800.

[24] Angela M. Hancock, Benjamin Brachi, Nathalie Faure, Matthew W. Horton, Lucien B. Jarymowycz, F. Gianluca Sperone, Chris Toomajian, Fabrice Roux, and Joy Bergelson. Adaptation to climate across the arabidopsis thaliana genome. *Science*, 334(6052):83–86, October 2011. ISSN 1095-9203. doi: 10.1126/science.1209244. URL http://dx.doi.org/10.1126/science.1209244.

[25] A. Fournier-Level, A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, and A. M. Wilczek. A map of local adaptation in arabidopsis thaliana. *Science*, 334(6052):86–89, October 2011. ISSN 1095-9203. doi: 10.1126/science.1209271. URL http://dx.doi.org/10.1126/science.1209271.

[26] JESSE R. LASKY, DAVID L. DES MARAIS, JOHN K. McKAY, JAMES H. RICHARDS, THOMAS E. JUENGER, and TIMOTHY H. KEITT. Characterizing genomic variation of arabidopsis thaliana : the roles of geography and climate. *Molecular Ecology*, 21(22):5512–5529, August 2012. ISSN 1365-294X. doi: 10.1111/j.1365-294x.2012.05709.x. URL http://dx.doi.org/10.1111/j.1365-294X.2012.05709.x.

[27] Moises Exposito-Alonso, François Vasseur, Wei Ding, George Wang, Hernán A. Burbano, and Detlef Weigel. Genomic basis and evolutionary potential for extreme drought adaptation in arabidopsis thaliana. *Nature Ecology amp; Evolution*, 2(2):352–358, December 2017. ISSN 2397-334X. doi: 10.1038/s41559-017-0423-0. URL http://dx.doi.org/10.1038/s41559-017-0423-0.

[28] Moises Exposito-Alonso, 500 Genomes Field Experiment Team, Hernán A. Burbano, Oliver Bossdorf, Rasmus Nielsen, and Detlef Weigel. Natural selection on the arabidopsis thaliana genome in present and future climates. *Nature*, 573(7772):126–129, August 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1520-9. URL http://dx.doi.org/10.1038/s41586-019-1520-9.

[29] Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904–909, 2006. ISSN 1546-1718. doi: 10.1038/ng1847. URL http://dx.doi.org/10.1038/ng1847.

[30] Jianming Yu, Gael Pressoir, William H Briggs, Irie Vroh Bi, Masanori Yamasaki, John F Doebley, Michael D McMullen, Brandon S Gaut, Dahlia M Nielsen, James B Holland, Stephen Kresovich, and Edward S Buckler. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2):203–208, December 2005. ISSN 1546-1718. doi: 10.1038/ng1702. URL http://dx.doi.org/10.1038/ng1702.

[31] Hyun Min Kang, Jae Hoon Sul, Susan K Service, Noah A Zaitlen, Sit-yee Kong, Nelson B Freimer, Chiara Sabatti, and Eleazar Eskin. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4):348–354, March 2010. ISSN 1546-1718. doi: 10.1038/ng.548. URL http://dx.doi.org/10.1038/ng.548.

[32] David L. Des Marais, Kyle M. Hernandez, and Thomas E. Juenger. Genotype-by-environment interaction and plasticity: Exploring genomic responses of plants to the abiotic environment. *Annual Review of Ecology, Evolution, and Systematics*, 44(1):5–29, November 2013. ISSN 1545-2069. doi: 10.1146/annurev-ecolsys-110512-135806. URL http://dx.doi.org/10.1146/annurev-ecolsys-110512-135806.

[33] Trudy F. C. Mackay, Eric A. Stone, and Julien F. Ayroles. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, 10(8):565–577, August 2009. ISSN 1471-0064. doi: 10.1038/nrg2612. URL http://dx.doi.org/10.1038/nrg2612.

[34] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I.W. de Bakker, Mark J. Daly, and Pak C. Sham. Plink: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3):559–575, 2007. ISSN 0002-9297. doi: 10.1086/519795. URL http://dx.doi.org/10.1086/519795.

[35] Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), February 2015. ISSN 2047-217X. doi: 10.1186/s13742-015-0047-8. URL http://dx.doi.org/10.1186/s13742-015-0047-8.

[36] Xiang Zhou and Matthew Stephens. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7):821–824, 2012. ISSN 1546-1718. doi: 10.1038/ng.2310. URL http://dx.doi.org/10.1038/ng.2310.

[37] Christoph Lippert, Jennifer Listgarten, Ying Liu, Carl M Kadie, Robert I Davidson, and David Heckerman. Fast linear mixed models for genome-wide association studies. *Nature*

Methods, 8(10):833–835, 2011. ISSN 1548-7105. doi: 10.1038/nmeth.1681. URL http://dx.doi.org/10.1038/nmeth.1681.

[38] Cristen J. Willer, Yun Li, and Gonçalo R. Abecasis. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191, 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq340. URL http://dx.doi.org/10.1093/bioinformatics/btq340.

[39] Julian P. T. Higgins and Simon G. Thompson. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11):1539–1558, May 2002. ISSN 1097-0258. doi: 10.1002/sim.1186. URL http://dx.doi.org/10.1002/sim.1186.

[40] B. Devlin and Kathryn Roeder. Genomic control for association studies. *Biometrics*, 55(4): 997–1004, December 1999. ISSN 1541-0420. doi: 10.1111/j.0006-341x.1999.00997.x. URL http://dx.doi.org/10.1111/j.0006-341x.1999.00997.x.

[41] Steven J. Franks, Sheina Sim, and Arthur E. Weis. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences*, 104(4):1278–1282, January 2007. ISSN 1091-6490. doi: 10.1073/pnas.0608379104. URL http://dx.doi.org/10.1073/pnas.0608379104.