Genomic architecture of seed weight and progeny germination under contrasting maternal environments in common bean

Martin Wohlfahrt

Degree project/Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Faculty of Natural Resources and Agricultural Sciences/ Department of Plant Biology
Plant Biology for Sustainable Production - Master´s programme - Abiotic and Biotic Interactions of Cultivated Plants - LM011
Uppsala 2025

Genomic architecture of seed weight and progeny germination under contrasting maternal environments in common bean.

Martin Wohlfahrt

Supervisor:	Martha Rendón Anaya, SLU, Department of Plant Biology
Assistant supervisor:	Pär Ingvarsson, SLU, Department of Plant Biology
Examiner:	Salim Bourras, SLU, Department of Plant Biology

Credits:	30 credits
Level:	Master
Course title:	Independent project in Biology (A2E)
Course code:	EX0856
Program:	Plant Biology for Sustainable Production - Master´s programme - Abiotic and Biotic Interactions of Cultivated Plants
Course coordinating dept:	Department of Plant Biology
Place of publication:	Uppsala
Year of publication:	2025

Swedish University of Agricultural Sciences
Faculty of Landscape Architecture, Horticulture and Crop Production Sciences
Department of Plant Biology

Abstract

Drought is a leading constraint to common bean (Phaseolus vulgaris L.) production, yet the genetic basis of drought-resilient seed weight and progeny vigor remains incompletely resolved. This study integrated phenotyping of seed weight (GW100) and germination traits with genome-wide association (GWAS), and candidate-gene curation in a structured panel. Seeds originated from plants grown under either terminal drought or well-watered conditions ("maternal environments"). In this 38-accession germination subset, maternal drought reduced seed mass by ~15.3% on average, with strategy-dependent magnitudes, whereas within-environment correlations between GW100 and germination index (GI), synchrony (T80T20), or final germination rate (GR) were small and non-significant; pooled models indicated no global maternal-environment × weight interaction, pointing to predominant genotype effects on germination dynamics.

Across the harmonized panel (n = 170), GWAS identified 13 lead SNPs spanning GW100 under reference and drought conditions, GI_drought, GR_drought, and T80T20, with nine genome-wide and four suggestive signals. LD inspection motivated uniform ±150 kb core windows for downstream interpretation. Candidate mining within these windows (Phaseolus vulgaris v2.1; Arabidopsis orthologs via UniProt) yielded >130 genes overall, of which 36 were stress-related (e.g., receptor-like kinases, redox enzymes, transcription factors), alongside genes implicated in seed/reproductive biology. Colocalization supported biological plausibility, including overlaps with classic seed-size QTL (SW2.1; SL8.1), drought-yield regions on Chr09, a domestication/shattering interval around PvPdh1, and seed-quality (cooking-time) QTL plausibly linked to seed-coat properties.

Collectively, the results indicate that maternal drought commonly reduces seed mass but that progeny germination responses are genotype-structured and not well predicted by seed weight alone. The mapped loci, compact LD-anchored candidate sets, and QTL overlaps provide tractable entry points for fine-mapping and validation toward breeding for drought-resilient seed quality and yield.

Keywords: Phaseolus vulgaris; drought; seed weight (GW100); germination; GWAS; linkage disequilibrium; candidate genes; QTL colocalization; maternal environment; seed vigor.

Table of contents

1. Introduction

The common bean (*Phaseolus vulgaris* L.) is a major staple for more than 300 million people due to its favorable nutritional profile, featuring high protein, soluble fiber, complex carbohydrates, and mineral density, alongside its adaptability across smallholder systems (Beebe et al., 2013; Smith et al., 2019). It can contribute up to one-third of daily protein intake, underscoring its role in food security (Schmutz et al., 2014). However, production is predominantly rainfed and thus vulnerable to episodic and terminal droughts that can incur severe yield penalties. Multi-environment reports attribute up to ~80% losses in extreme seasons, placing drought among the leading causes of yield failure after disease (Rosales-Serna et al., 2004; Villordo-Pineda et al., 2015). Physiological responses to water deficit are commonly framed as four strategies described in a consistent manner: tolerance, the capacity to maintain function at low tissue water potential via osmotic adjustment and cell-wall elasticity; avoidance, the maintenance of plant water status through rooting depth, stomatal regulation, and water conservation; escape, the acceleration of phenology and remobilization to complete reproduction before severe deficit; and recovery, the ability to re-green and resume growth after re-watering (Beebe et al., 2013; Rosales-Serna et al., 2004). In addition, a stay-green (SG) strategy denotes delayed foliar senescence that sustains photosynthetic capacity during stress; SG occurs in two forms, functional SG (photosynthesis maintained) and cosmetic SG (chlorophyll retained but photosynthetic competence lost), with functional SG further observed as Type A (delayed onset of senescence) or Type B (normal onset with slower progression) (Thomas and Ougham, 2014; Kamal et al., 2019).

Seed qualities and yield output are quantitatively inherited and strongly context-dependent in common bean. Classical QTL studies resolved loci for seed size and related yield components across several chromosomes, while diversity-panel GWAS extended this picture to many small-effect, environment-sensitive associations (Tar'an et al., 2002; Blair et al., 2006; Moghaddam et al., 2016). Under stress, dedicated mapping has identified drought-yield QTL and indicated possible linkage/pleiotropy with domestication and dehiscence loci such as *PvPdh*1, suggesting shared genetic neighborhoods among adaptation traits, pod shattering, and yield stability (Trapp et al., 2015; Parker et al., 2020; Pour-Aboughadareh et al., 2022; Blair et al., 2010). Together, these studies support a polygenic model with partially distinct and partially overlapping architectures under well-watered versus water-limited conditions.

Recent phenotyping of natural variation has begun to link whole-plant drought strategies to reproductive success. For example, Labastida et al. (2023) surveyed multiple gene pools, highlighted accessions with robust stay-green behavior, and nominated candidate genes consistent with delayed senescence and sustained assimilate

supply into seed filling. In that study, accessions were operationally assigned to stay-green (greenness maintained in stems and leaves during stress), escape (increased pod set under drought with <75% yield loss), recovery (re-greening and renewed leaf/pod production after re-watering), or susceptible (>75% yield loss or plant death) (Labastida et al., 2023). Such strategy-level differences provide testable hypotheses about how genotypes partition yield resilience between source maintenance and sink remobilization under terminal stress.

Against this backdrop, this study targets the genomic architecture of seed traits (such as seed weight or germination rates) from seeds of two different maternal backgrounds: Seeds produced by plants subjected to terminal drought for two weeks before re-watering and plants kept under optimum (well-watered) conditions. Specific objectives are: (i) to quantify condition-specific and shared loci influencing grain yield across contrasting water regimes; (ii) to benchmark signals against established QTL for seed size, seed-quality, and agronomic performance to assess biological plausibility and potential trade-offs; (iii) to integrate population structure and drought-response strategy as covariates for improved calibration; and (iv) to prioritize candidate genes within local LD windows for downstream validation. Additionally, (v) generating a comprehensive reference list for seed weight (100 grain weight) for many European accessions. Framing yield genetics across environments aims to inform breeding for drought-resilient yield, complementing prior emphasis on yield potential under optimum management (White et al., 1994; Tar'an et al., 2002; Blair et al., 2006; Trapp et al., 2015; Pour-Aboughadareh et al., 2022; Labastida et al., 2023). Climate-driven increases in the frequency and intensity of heat waves and terminal droughts are constraining seed production and seedlot quality, elevating the importance of seed vigor as a predictor of stand establishment under variable field conditions. Because germination capacity and vigor are acquired during seed development, the maternal environment during flowering, seed filling, and maturation can modulate dormancy, longevity, and germination kinetics, with consequences for crop establishment under stress (Brunel-Muguet et al., 2025).

Maternal stress memory provides a mechanistic and applied framework for these effects: stress exposures before or after fertilization can leave inter-, intra-, or transgenerational imprints—often via epigenetic marks and small RNAs—that influence germination and early seedling performance. A recent Plant Journal viewpoint outlines a roadmap for “climate-smart” seedlots by harnessing maternal priming, while emphasizing that the stability and predictability of such imprints require further validation (Brunel-Muguet et al., 2025). In this context, comparing seeds produced under terminal drought versus well-watered conditions directly tests whether maternal environments contribute to seed-vigor differences in common bean and helps identify

candidate mechanisms and loci that could be leveraged to buffer climate risk in seed systems.

2. Material and Methods

2.1 Plant Material

The experimental material consisted of common bean (*Phaseolus vulgaris* L.) seeds derived from a previous drought experiment (Labastida et al. 2023), in which plants representing 71 accessions from five gene pools—domesticated Andean (A), domesticated Mesoamerican (MA), European (EU), Andean wild (AW), and Mesoamerican wild (MW)—were subjected to terminal drought stress or maintained under well-watered conditions. Plant material was provided by the International Center for Tropical Agriculture (CIAT), the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, and the Nordic Genetic Resource Center (NordGen). Seeds harvested from drought-stressed and control plants were used for the present germination assay; due to low seed set in drought-susceptible accessions, the final germination subset comprised 38 accessions, primarily of European origin. In parallel, a broader reference panel was assembled to characterize seed mass by gene pool: GW100 (g per 100 seeds) measured on non-drought reference harvests was compiled for 170 accessions. European reference seed weights were measured directly, whereas Central American/Mesoamerican values were obtained from the SEAD database. This GW100 dataset was used for gene-pool comparisons downstream (Fig. 1 and 10), while analyses by maternal environment (control vs terminal drought) were conducted on the 38-accession germination subset. Per-group sample sizes were EU = 98, AW = 21, MW = 20, MA = 16, and A = 15. The full accession list can be viewed in Table 5 (Appendix) .

2.2 Seed weight measurements

Prior to the germination assay, seed weight was determined for both control-derived seeds and drought-derived seeds. Hundred-seed weight (GW100) was calculated separately for each maternal environment (control vs. drought). Additionally, seed weight data (GW100, g) was obtained for 170 accessions. European accessions were measured directly from reference seed material, as sufficient data was not available from the resources of Nordgen and IPK, whereas values for Andean, Mesoamerican, and wild gene pools were retrieved from publicly available data provided by the Alliance of Bioversity International and CIAT (part of CGIAR).

2.3 Germination experiment

Seeds were obtained exclusively from the prior drought experiment. For each of the 38 accessions, ten seeds (five produced under terminal drought and five produced under well-watered conditions) were placed on moistened paper in sterile Petri dishes, wetted

with 40 mL tap water, and maintained at ≈21–23 °C. Germination was recorded every 24 h for 7 days; a seed was scored as germinated when the radicle exceeded 1 mm.

2.4 Germination traits

The following germination parameters were determined:

- Germination Index (GI):

  The GI was calculated according to the Association of Official Seed Analysts (AOSA, 1983) records:

$$GI = \sum \left(\frac{GT}{T} \right)$$

  where GT is the number of seeds germinated on day T (not cumulative). Higher values indicate faster early germination or:

$$GI = \left(\frac{Number\ of\ germinated\ seeds\ in\ first\ count}{Day\ of\ first\ count} \right) + \cdots + \left(\frac{Number\ of\ germinated\ seeds\ in\ final\ count}{Day\ of\ final\ count} \right)$$

- T80T20: The interval (days) between 20% and 80% germination, used as a measure of germination synchrony.
- GR: Final germination percentage of germinated seeds at the end of the 7-day assay.

2.5 Statistical analysis of GW100 by gene pool and maternal environment effects

Metadata (accession ID, gene pool) and reference seed mass (GW100, g per 100 seeds) were imported in R (RStudio) using readxl and merged by accession; data wrangling used dplyr/tidyr with factor handling via forcats, and graphics via ggplot2. Exploratory visualization comprised one-bar-per-accession columns (X = accession, Y = GW100; colored by gene pool) and box-and-jitter plots of GW100 by gene pool. For inference on GW100 across gene pools (A, MA, EU, AW, MW), a one-way ANOVA (stats::aov) was inspected with Shapiro–Wilk residual normality (stats::shapiro.test) and Bartlett homogeneity (stats::bartlett.test); because assumptions were violated, group differences were evaluated using the Kruskal–Wallis rank-sum test (stats::kruskal.test) with epsilon-squared (ε^2) as effect size, followed by Wilcoxon rank-sum post-hoc contrasts with Holm

multiplicity adjustment (stats::pairwise.wilcox.test, p.adjust.method="holm"). Maternal-environment effects were quantified per accession as derived contrasts (drought − control) for GW100, germination index (GI), germination spread (T80T20), and final percentage of germinated seeds (GR); a weight-preservation ratio (drought GW100 / control GW100) provided a scale-free summary of seed-mass maintenance under maternal drought. Within each maternal environment (control, drought), monotonic associations between GW100 and germination traits were assessed using Spearman rank correlations (stats::cor.test, method="spearman"). Drought sensitivity (per-accession change scores) was modeled with ordinary least squares (stats::lm) as a function of baseline control-origin GW100 and, in a complementary specification, of the weight-preservation ratio, including predefined drought-response strategy classes (Escape, Recovery, Susceptible, Stay-Green, Not available) as covariates. To test maternal environment-specific weight effects, pooled linear models were fit with trait values as the response and maternal environment, environment-matched seed weight, strategy, and the environment×weight interaction as predictors. Distributions of change scores were summarized by location, spread, and counts of positive vs. negative values; model assumptions were checked by standard diagnostics (residual patterns, variance homogeneity). All tests were two-sided with $\alpha = 0.05$; no multiplicity adjustment was applied to the maternal-environment models, and multiplicity control was applied only to pairwise gene-pool contrasts. Visualization additionally included cumulative germination curves across days 1–7 with two lines per accession (one per maternal environment) faceted by strategy.

2.6 Principal component analyses

Trait PCA. Principal component analysis was conducted in R using stats::prcomp with centering and unit-variance scaling (i.e., PCA of the correlation matrix). The input comprised standardized phenotypes GI, T80T20, GR, and GW100 (where available); rows with missing values were excluded (complete-case analysis). Eigenvalues, variance explained, and loadings were extracted to summarize trait interrelationships. Concordance was verified with FactoMineR::PCA under identical settings.

Genetic PCA. Single-nucleotide polymorphism (SNP) data were used to assess population structure and to provide covariates for genome-wide association analysis. Principal component analysis (PCA) was carried out using PLINK v1.9, and the first five principal components (PCs) were retained. These components captured the principal axes of genetic variation and were subsequently included as covariates in the association models to correct for population stratification. Analyses operated on an upstream quality-filtered, LD-pruned marker set; no additional pruning was applied at the PCA stage. PLINK computed principal components by singular-value decomposition of the standardized genotype matrix (each SNP centered at 2p and scaled by $\sqrt{[2p(1-p)]}$), using

mean imputation for missing genotypes. The resulting eigenvectors also provided a low-dimensional summary for visualization of population structure.

Genotype dataset summary. The genome-wide catalog contained 161,391 SNPs (SNP-only; indels excluded), distributed across chromosomes as follows: Chr01, 15,729; Chr02, 15,248; Chr03, 15,737; Chr04, 17,667; Chr05, 12,969; Chr06, 9,128; Chr07, 11,291; Chr08, 21,149; Chr09, 10,607; Chr10, 13,827; Chr11, 18,039. Chromosome 8 contributed the largest share, whereas chromosome 6 contributed the fewest. PCA operated on the previously LD-pruned subset of this catalog, while the full distribution provides context for genome coverage. Per-SNP MAF was summarized for descriptive purposes and allele-frequency stratification; unless otherwise stated, no MAF exclusion threshold was applied in the primary analyses.

2.7 Genome-wide association study (GWAS)

Genome-wide association analyses were performed in R 4.4.1 (RStudio 2023.9.1.494) using GAPIT3 v3.5.0 (Lipka et al., 2012), with BLINK specified as the sole model for the final scans (Huang et al., 2019). Models included an intercept and PC1–PC5 as fixed covariates and were run on the harmonized panel (n = 170) and the filtered biallelic SNP set described above; during method development, alternative GWAS frameworks—MLM (Zhang et al., 2010), MLMM (Segura et al., 2012), and FarmCPU (Liu et al., 2016)—were considered under identical quality-control and covariate settings, with a VanRaden genomic relationship matrix applied where appropriate (VanRaden, 2008). Model choice was based on calibration diagnostics: comparative runs under identical QC and covariates indicated that BLINK produced the best-aligned quantile–quantile plots and genomic inflation factors closest to 1, and was therefore retained for inference. Accordingly, only BLINK results are reported. Associations were declared genome-wide significant at a Bonferroni threshold (α=0.05; $p < 3.10\times10^{-7}$ for m=161,391 tests). Suggestive associations were flagged at $p < 1/m = 6.20\times10^{-6}$. Benjamini–Hochberg FDR q-values were computed for reference, but were not used to define significance by GAPIT to visualize association signals and evaluate model fit.

2.8 Phenotype imputation and zero-coding rationale

For GW100_ref, 170 accessions had observed values and were analyzed directly. In case of traits with limited observed sample sizes (GW100 under drought, and the germination-derived traits), for GWAS sensitivity only, auxiliary zero-coded vectors (unmeasured accessions set to 0) were created to preserve genome-wide coverage in BLINK. These zero-coded analyses are reported exclusively as sensitivity checks in the Supplementary and did not replace the primary inferences based on observed data, given their potential to induce zero-inflation and attenuate effect-size interpretability.

2.9 Linkage disequilibrium (LD) window calculations

The SNPs obtained from the prior performed GWAS served as focal sites for linkage disequilibrium (LD) analyses. Local LD was computed and visualized with LDBlockShow using genotype calls from the variant dataset. For each index SNP, two symmetric windows were interrogated: a close-up region (±50 kb) and an overview region (±500 kb). Because pairwise LD estimates from the cohort did not support stable numeric thresholds across loci, LD bounds were delineated by visual inspection of the D′ heatmaps generated by LDBlockShow at both scales. For each locus, the final LD interval was called at the first consistent transition from densely high-D′ tiles to a fragmented signal visible in both panels.

2.10 Gene mapping and functional annotation

Candidate genes were retrieved within the ±150 kb core LD window centered on each lead SNP, as defined by visual inspection of D′ heatmaps. This uniform window ensures comparable locus sizes across traits; any locus-specific deviations are stated explicitly in the results. All annotated genes falling within the aforementioned intervals were retrieved from the *P. vulgaris* reference genome (Phytozome v2.1). Gene functions were assigned using the functional descriptions available in the reference annotation. To further refine the annotations, putative orthologs in *Arabidopsis thaliana* were identified, and their functions were confirmed using curated records in UniProt. On this basis, each gene was classified according to its likely biological role, with particular attention given to categories related to abiotic stress responses and seed or reproductive development.

2.11 Overlap with published QTL/GWAS intervals

Chromosomal positions and trait associations for each SNP were cross-referenced with peer-reviewed QTL and GWAS literature to evaluate potential colocalization with previously reported loci in *P. vulgaris*. A QTL was considered co-localized if the physical position of the SNP fell within or in close proximity (≤250 kb) to the reported boundaries of the QTL. The window of ±250 kb was applied in consistency with the extent of linkage disequilibrium (LD) decay observed in *P. vulgaris* (Ariani et al., 2018), and in line with thresholds commonly adopted in recent GWAS–QTL integration studies in legumes (Díaz et al., 2020; Li et al., 2020). Where applicable, the associated trait and full citation of the corresponding publication were recorded.

3. Results

3.1 Seed Weight

Across the 38-accession germination subset, maternal terminal drought reduced reference seed mass by ≈5.37 g per 100 seeds (~15.3%; control − drought), with the largest absolute and proportional losses in Escape and Stay-green, intermediate losses in Recovery and NA, and the smallest reduction in Susceptible (Table 1). Extending to the full reference panel (n = 170), GW100 spanned a wide range from low single-digit values to near 100 g per 100 seeds, and clear among-pool differences were evident: means (g per 100 seeds) were A = 47.95, EU = 42.94, MA = 32.42, AW = 10.52, and MW = 7.15, yielding the ordered pattern MW < AW < MA < EU < A. Assumption checks supported the use of distribution-free methods: Shapiro–Wilk indicated non-normality ($W = 0.945$, $p = 3.42 \times 10^{-6}$) and Bartlett's test indicated heteroscedasticity across pools ($K^2 = 100.43$, $df = 4$, $p < 2.2 \times 10^{-16}$). A global Kruskal–Wallis test then rejected equal distributions among pools ($\chi^2 = 96.783$, $df = 4$, $p < 2.2 \times 10^{-16}$) with a large effect size ($\varepsilon^2 = 0.562$), and Holm-adjusted Wilcoxon contrasts localized the differences: both wild pools (AW, MW) were significantly lighter than A, EU, and MA, whereas contrasts among A, EU, and MA were not significant after adjustment. A Welch ANOVA provided a parametric robustness check that aligned with the rank-based inference ($F = 116.27$, $df = 4, 45.684$, $p < 2.2 \times 10^{-16}$). Collectively, the descriptive means, the violation of parametric assumptions, and the convergent results from non-parametric and Welch tests demonstrate pronounced gene-pool structuring of seed mass, with domesticated pools carrying substantially heavier seeds than the wild pools, and MA occupying an intermediate position (Table 1; Figure 1)

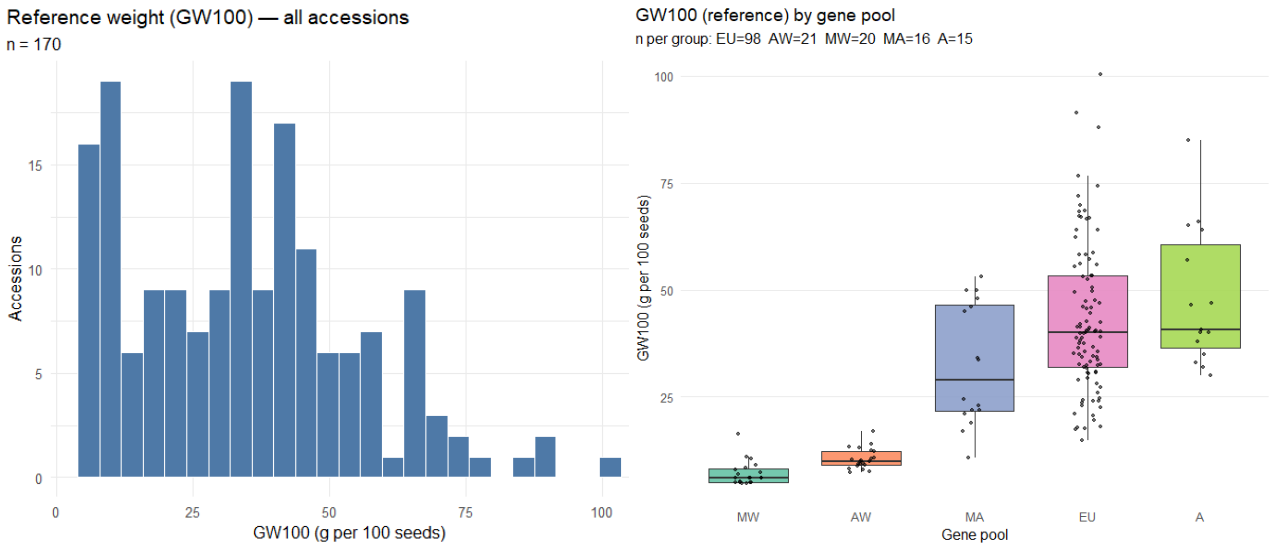


Figure 1: Reference GW100 distribution (left) and GW100 by gene pool (right) box-and-jitter per genetic group.

Strategy	Control mean (g)	Drought mean (g)	Difference (g)	Change (%)
Overall (n=38)	35.08	29.71	5.37	15.3
Escape (n=11)	35.36	28.79	6.58	18.6
Recovery (n=9)	37.07	31.78	5.28	14.3
Susceptible (n=10)	35.64	31.23	4.41	12.4
Stay-green (n=3)	33.91	27.91	6.00	17.7
NA (n=5)	30.43	26.01	4.42	14.5

Table 1: Strategy-wise GW100 means under control and drought and their differences. 'n' refferes to the number of accessions.

3.2 Influence of maternal environment and seed weight

Although seeds produced under maternal terminal drought were typically lighter (Table 1), GW100 showed no clear association with germination performance. Within the control environment, Spearman correlations between GW100 and GI, T80–T20, and GR were −0.012, −0.009, and 0.083, respectively; within the drought environment, the corresponding ρ values were −0.198, 0.053, and −0.079, all small and non-significant. Change-score regressions of ΔTrait (drought − control) on baseline control weight (wC) yielded slopes that were statistically indistinguishable from zero (GI: −0.008, $p = 0.513$; T80–T20: 0.013, $p = 0.123$; GR: −0.316, $p = 0.288$), and environment × weight interaction terms were likewise non-significant (GI: $p = 0.436$; T80–T20: $p = 0.484$; GR: $p = 0.208$). These results indicate that, at the panel level, neither seed weight nor its interaction with maternal environment explains variation in germination traits. Instead, responses were structured primarily by genotype: escape accessions germinated robustly irrespective of maternal environment; recovery accessions showed intermediate and variable responses; susceptible accessions were most consistently reduced under drought; and stay-green accessions exhibited mixed effects, including both negative and positive responses. Collectively, the correlation, regression, and interaction estimates support genotype as the dominant driver of germination dynamics, with maternal drought and seed size exerting limited influence at the global scale.

3.3 Germination performance across drought strategies

Cumulative germination curves were calculated and revealed distinct patterns among the four drought response strategies (Escape, Recovery, Susceptible, and Stay-Green). Although first-day emergence was higher in drought-origin seeds, cumulative curves and timing metrics indicate overall slower germination (Fig. 2)

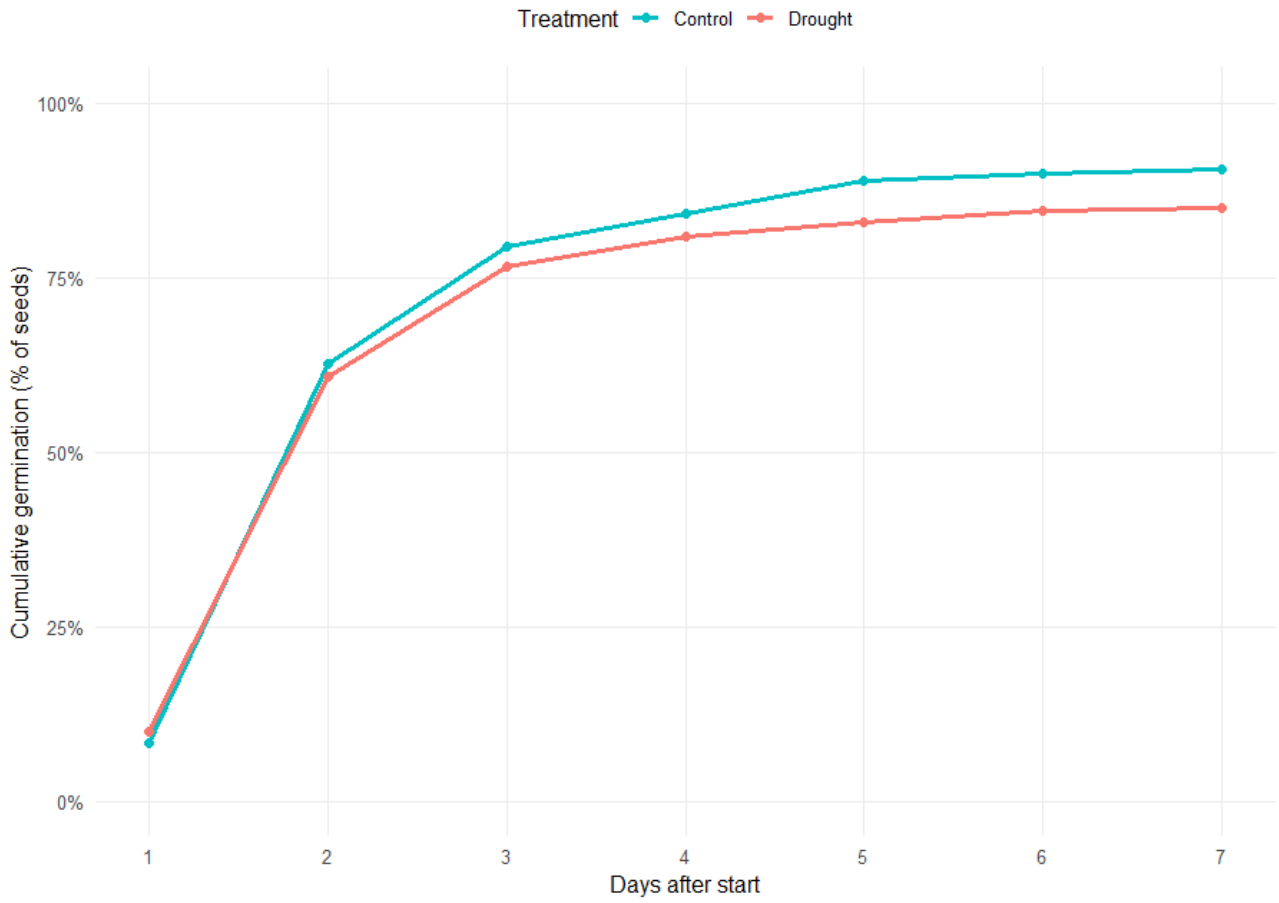


Figure 2: Percentual cumulative germination over time by treatment (c = control, d = maternal drought).

Escape accessions generally germinated rapidly and reached high final percentages (>75–100%) within the first 2–3 days. Maternal drought stress had little impact on their performance, with some accessions (e.g., G14629, PHA13928) even germinating slightly faster when derived from drought-stressed plants. Only a few accessions (e.g., PHA13666, PHA49) showed consistently lower maximum germination, but this appeared to be genotype-intrinsic rather than treatment-related. (Fig. 3)

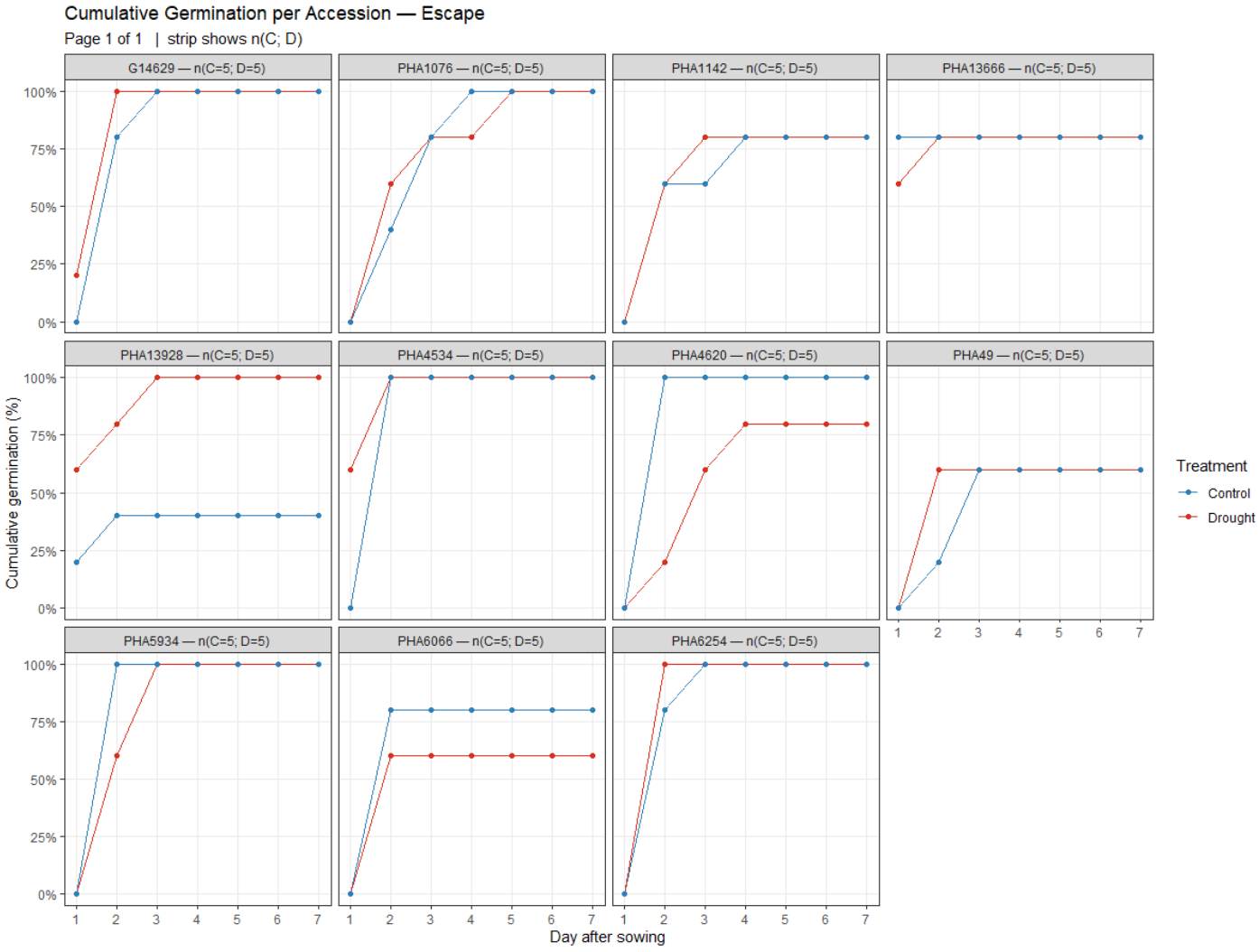


Figure 3: Cumulative germination (%) across days for Escape-class accessions

Recovery accessions showed more variable responses. While some lines (PHA1139, PHA13035, PHA4637) maintained high germination rates irrespective of treatment, others (e.g., G3296, PHA13099, PHA5989) exhibited clear reductions under drought, in some cases dropping from 100% in controls to below 50%. In several cases (e.g., G7930), drought-derived seeds germinated earlier than controls but failed to reach the same final percentages, suggesting a trade-off between initiation speed and overall germination success. (Fig. 4)

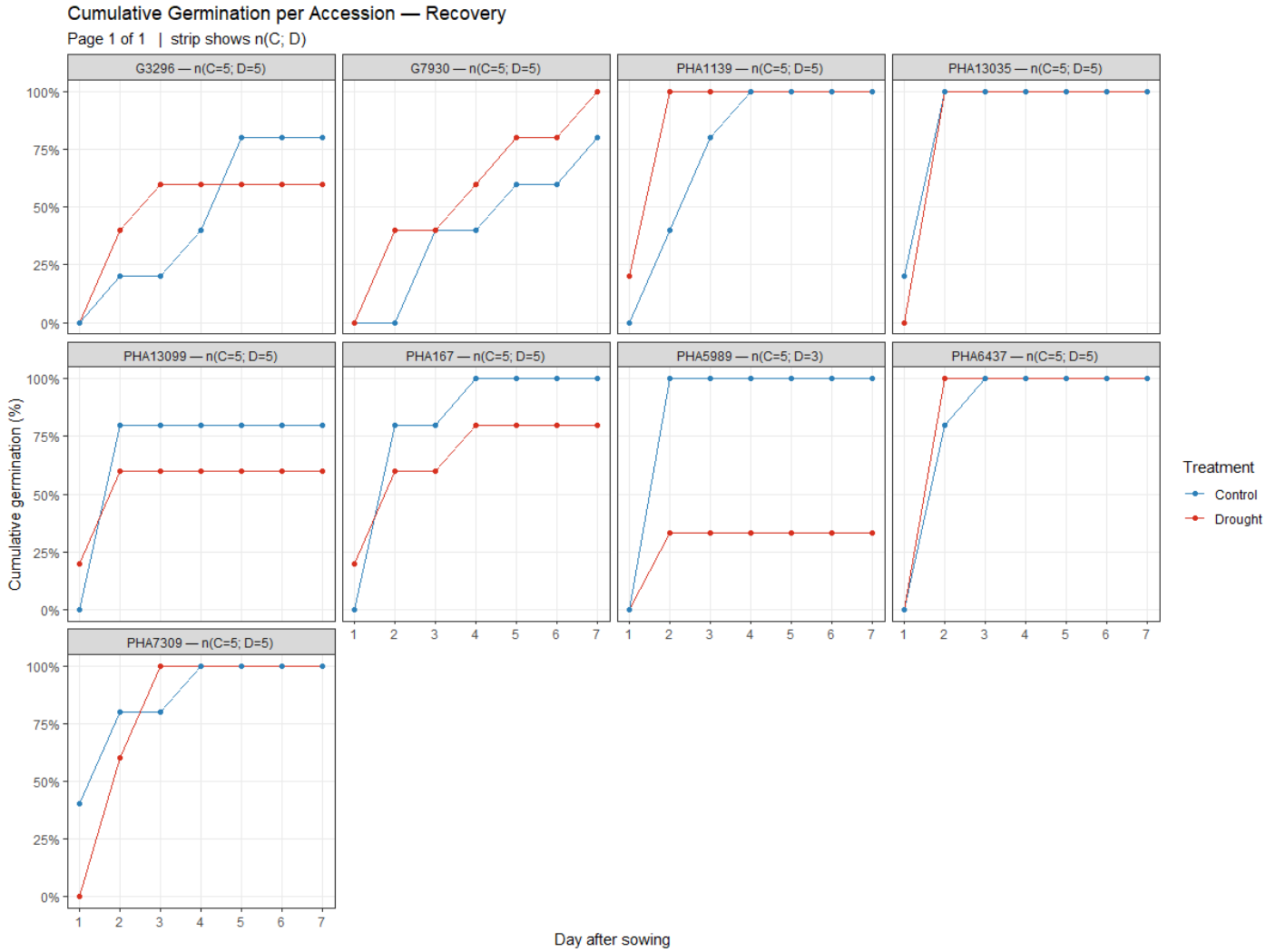


Figure 4: Cumulative germination (%) across days for Recovery-class accessions

Susceptible accessions were most strongly affected by maternal drought stress. Most susceptible lines could not be evaluated for germination because they failed to produce seed after the drought treatment, and thus were excluded from maternal-environment comparisons. Several lines (PHA14278, PHA1772, PHA6011) displayed markedly reduced final germination percentages (~20–70% under drought versus ~100% in controls) and delayed germination onset, indicating impaired seed vigor. While a few accessions (PHA13736, PHA13960) reached similar levels across treatments, the group as a whole showed the strongest and most consistent negative drought effects. (Fig. 5)

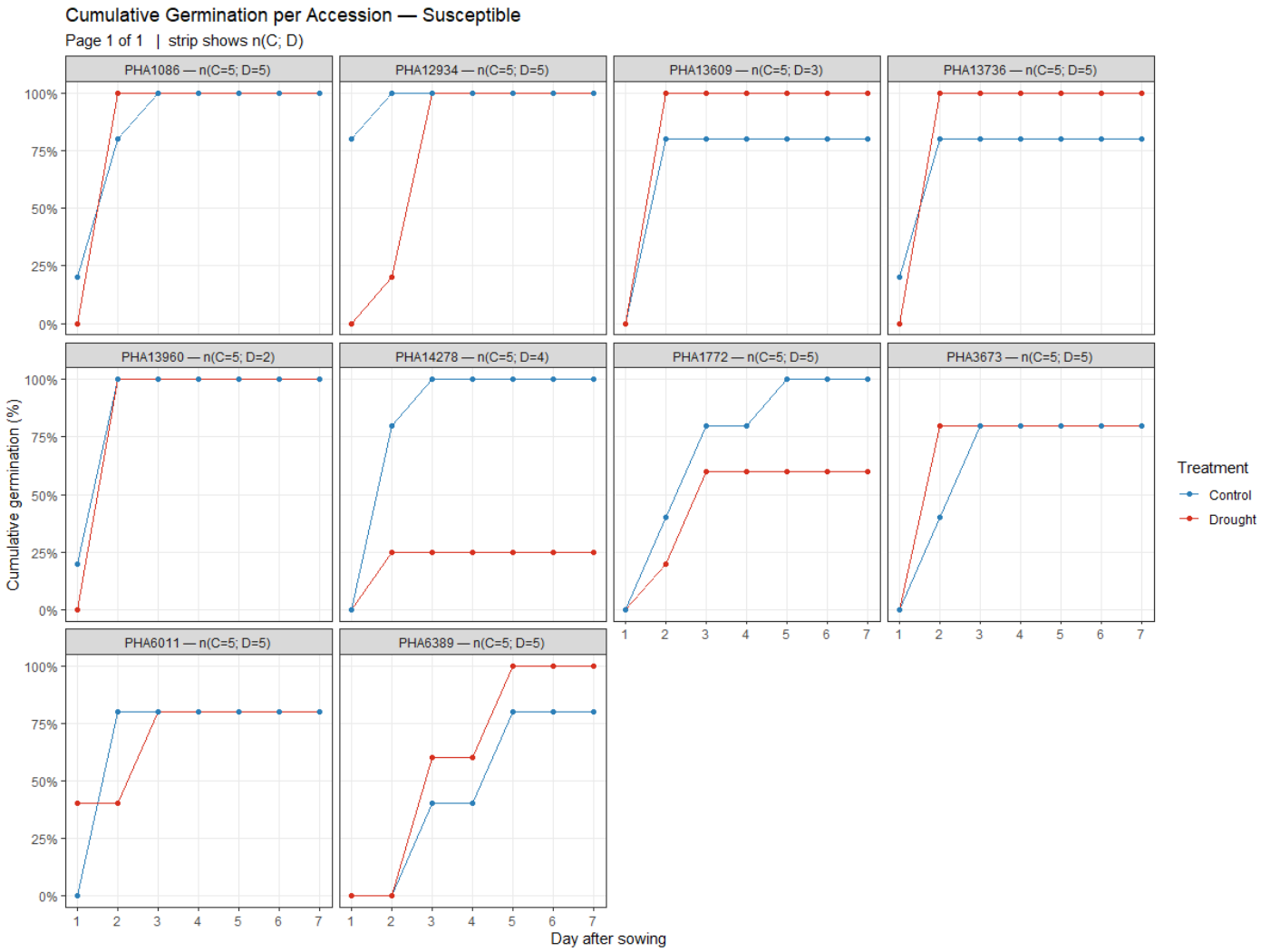


Figure 5: Cumulative germination (%) across days for Susceptible-class accessions

Stay-green accessions displayed mixed outcomes. Two accessions (PHA1077, PHA2682) exhibited reduced germination under drought (~75–80% vs. 100% in controls), while one accession (PHA6155) showed the opposite trend, with drought-derived seeds outperforming controls (100% vs. ~60%). Across the group, germination occurred rapidly within 2–3 days, and treatment effects were primarily reflected in final germination percentages rather than speed. (Fig. 6)

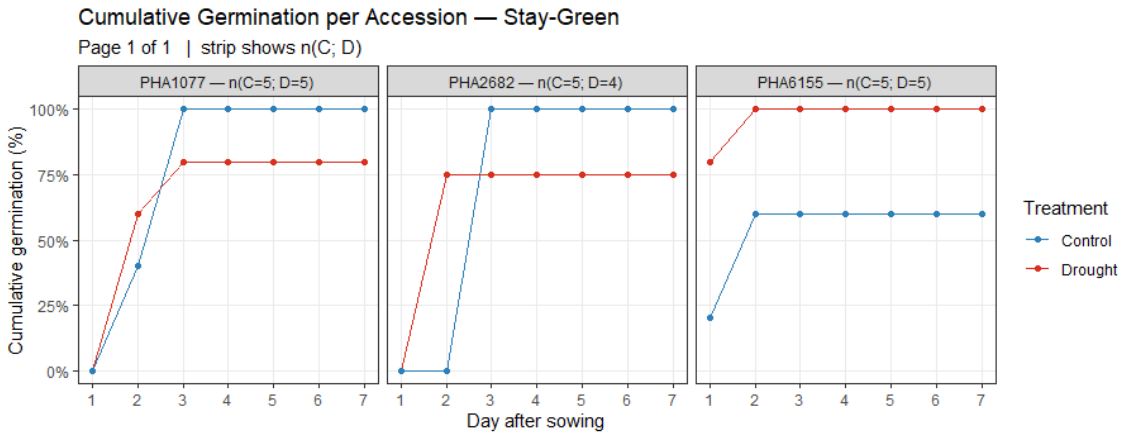


Figure 6: Cumulative germination (%) across days for Stay-green-class accessions

3.4 Phenotype-based PCA

Principal component analysis (PCA) integrating seed weight, yield-related, and germination traits explained 65% of the observed variance, with PC1 (47%) driven mainly by seed weight (GW100) and related yield traits, while PC2 (18%) was associated with germination dynamics (GI, GR, and germination homogeneity) (Fig. 7). This separation indicates that seed size and germination vigor represent related but distinct axes of variation. Genotypic drought-response strategies showed partial structuring in multivariate space, with recovery types clustering toward lighter seed weights, escape types distributed across both principal components, and stay-green accessions remaining near the center. Several outliers, particularly Mesoamerican accessions (e.g., G3296, G12875, G23458), displayed unusual germination behavior relative to seed weight. Trait correlation analysis supported these patterns, revealing directionally positive but weak and mostly non-significant associations between seed weight and germination indices under both control and drought conditions (Fig. 7). For instance, GW100 correlated positively with GI and GR, indicating that larger seeds generally germinated faster and more uniformly. Negative correlations were observed between germination indices and

time-to-germination traits (T80T20), consistent with expectations that faster germination reduces lag and spread. Importantly, these relationships were significant under both control- and drought-produced seed conditions, although correlation strengths varied. Together, these results highlight that while drought stress reduced overall germination performance, genotypic differences in seed weight and stress strategy contributed substantially to variation in progeny seed vigor.

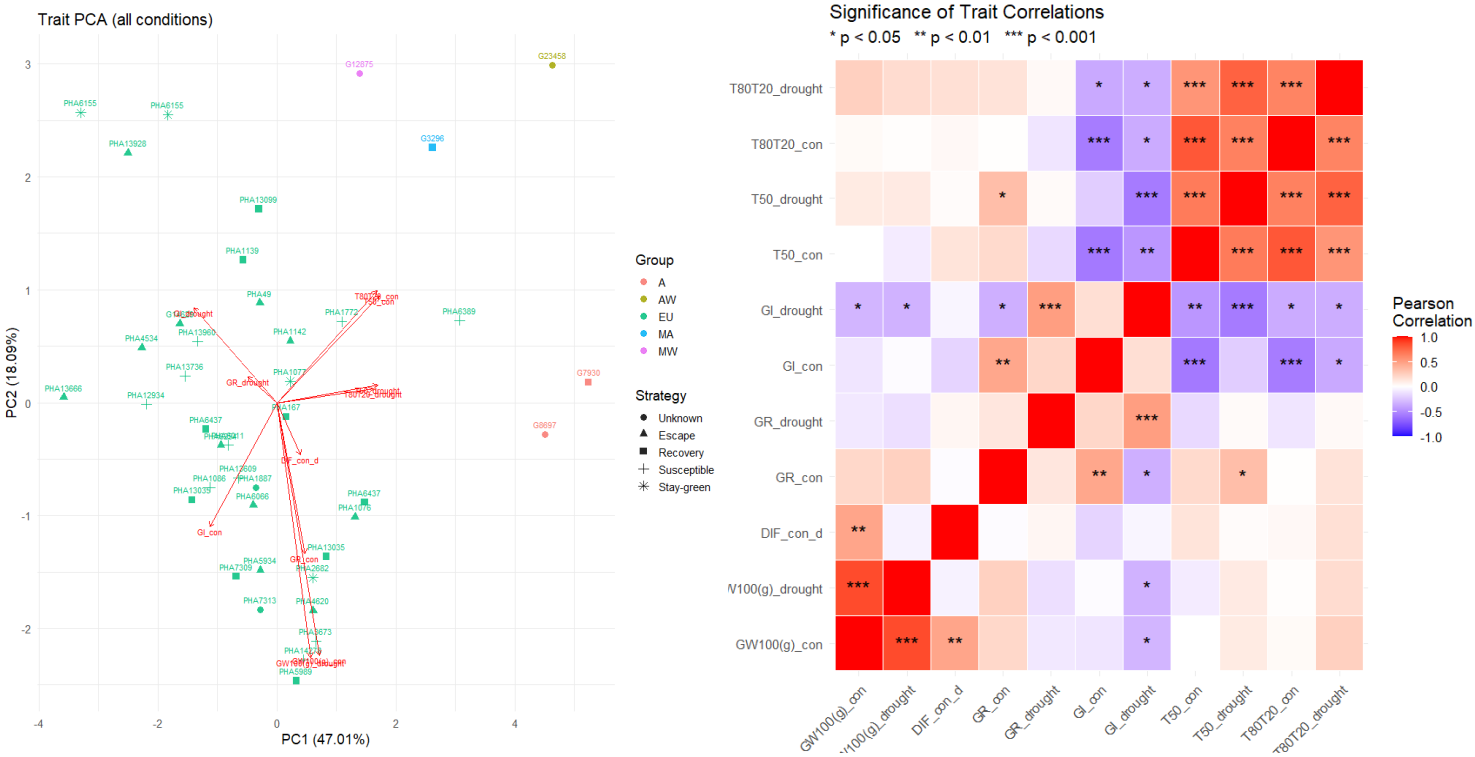


Figure 7: Multivariate structure of traits and their correlations. Left: PCA biplot of accessions using seed mass (GW100), yield-related, and germination traits from both seed origins. (drought and control). Points are colored by gene pool and shaped by drought-response strategy; arrows indicate trait loadings; axis labels show variance explained. Right: Pearson correlation matrix; asterisks denote significance (p < 0.05, ** p < 0.01, *** p < 0.001).*

3.5 SNP-based PCA

Population structure was assessed using principal component analysis (PCA) based on genome-wide SNP data. The full dataset comprised 170 genotypes, and the first five principal components were retained as covariates for GWAS to account for population stratification. An extract of this structure, focusing on the 38 genotypes that produced sufficient seed for the germination experiment, is shown in Fig. 8. The first two principal components explained 47.0% and 18.9% of the total variance, respectively. Accessions

grouped largely according to their genetic background, with European (EU) lines forming a compact cluster, while Mesoamerican (MA, MW) and Andean (A, AW) accessions appeared more dispersed. Notably, the domesticated Andean accessions G7930 and G8697 clustered proximal to the EU group, consistent with ancestral background: European common-bean germplasm descends from post-Columbian introductions from both domesticated gene pools, with a documented predominance of Andean ancestry and extensive introgression (Angioi et al., 2010; Pipan et al., 2019; Bellucci et al., 2023). Drought-response strategies were distributed across the genetic background rather than forming discrete clusters, although susceptible types were common within the European cluster, whereas stay-green and escape strategies were more broadly distributed. These results confirm the presence of clear population structure in the panel and justify the use of PCA covariates in GWAS to reduce spurious associations.

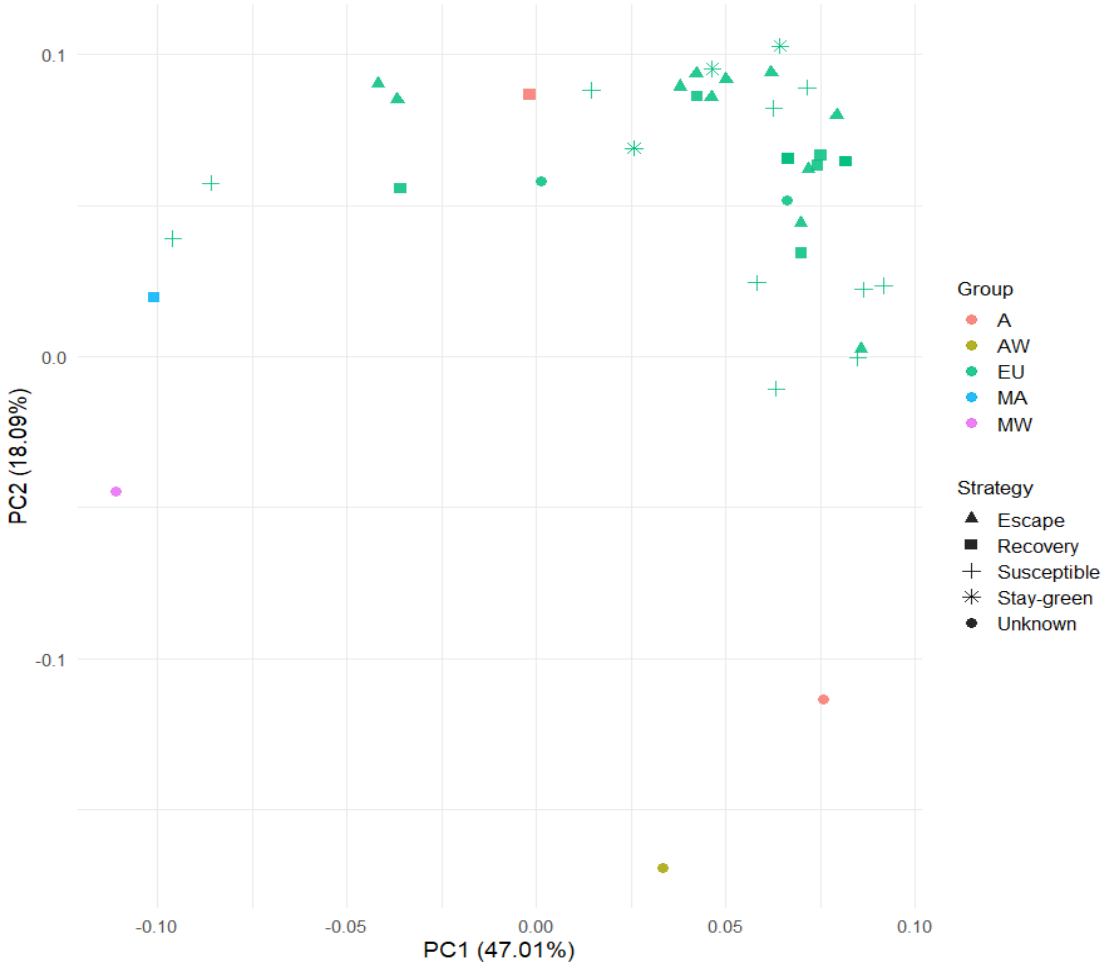


Figure 8: Population structure of the panel. PCA of genome-wide SNPs for the 38-accession germination subset. Axes show PC1 (47.01%) and PC2 (18.09%). Points are colored by gene pool (A, AW, EU, MA, MW) and shaped by drought-response strategy. European lines form a compact cluster, whereas Andean and Mesoamerican accessions are more dispersed; strategies are interspersed across genetic backgrounds.

3.6 GWAS

Genome-wide association across five traits (GW100_ref, GW100_drought, GI_drought, GR_drought, and T80T20) identified 13 lead SNPs in the harmonized panel (n = 170) using BLINK with PC1–PC5 as covariates. Calibration diagnostics (quantile–quantile behavior and genomic inflation) under identical quality control and covariates indicated that BLINK showed the least deviation from expectation and was consequently, retained for inference across traits (Fig. 9). For GW100_ref, top loci were on Chr8 and Chr2; for GW100_drought, two adjacent loci were on Chr9; for GI_drought, five loci were on Chr1, Chr3, Chr8, Chr10, and Chr7; for GR_drought, loci were on Chr8 and Chr11; and for T80T20, two independent peaks were on Chr7. Nine loci met the genome-wide threshold, and four were retained as suggestive (Table 3).

Trait	SNP	Chr	Pos	P.value
GI_drought	Chr01_15505693_C_T	1	15505693	4.875888e-10
	Chr03_9548597_G_A	3	9548597	1.547379e-08
	Chr07_22968375_A_C	7	22968375	1.588901e-07
	Chr08_33781184_C_T	8	33781184	1.945306e-08
	Chr10_8792003_G_A	10	8792003	7.815991e-08
GR_drought	Chr08_62503784_G_A	8	62503784	3.867279e-08
	Chr11_13009550_C_T	11	13009550	5.260349e-08
GW100_drought	Chr09_35392216_T_C	9	35392216	1.301392e-18
	Chr09_35434904_G_T	9	35434904	1.460840e-12
GW100_ref	Chr02_40631423_A_G	2	40631423	5.554790e-08
	Chr08_5236332_G_A	8	5236332	1.755408e-12
T80T20_drought	Chr07_606868_G_A	7	606868	1.244707e-09
	Chr07_6740141_C_T	7	6740141	3.118105e-08

Table 3: Lead SNPs from genome-wide association across five traits.

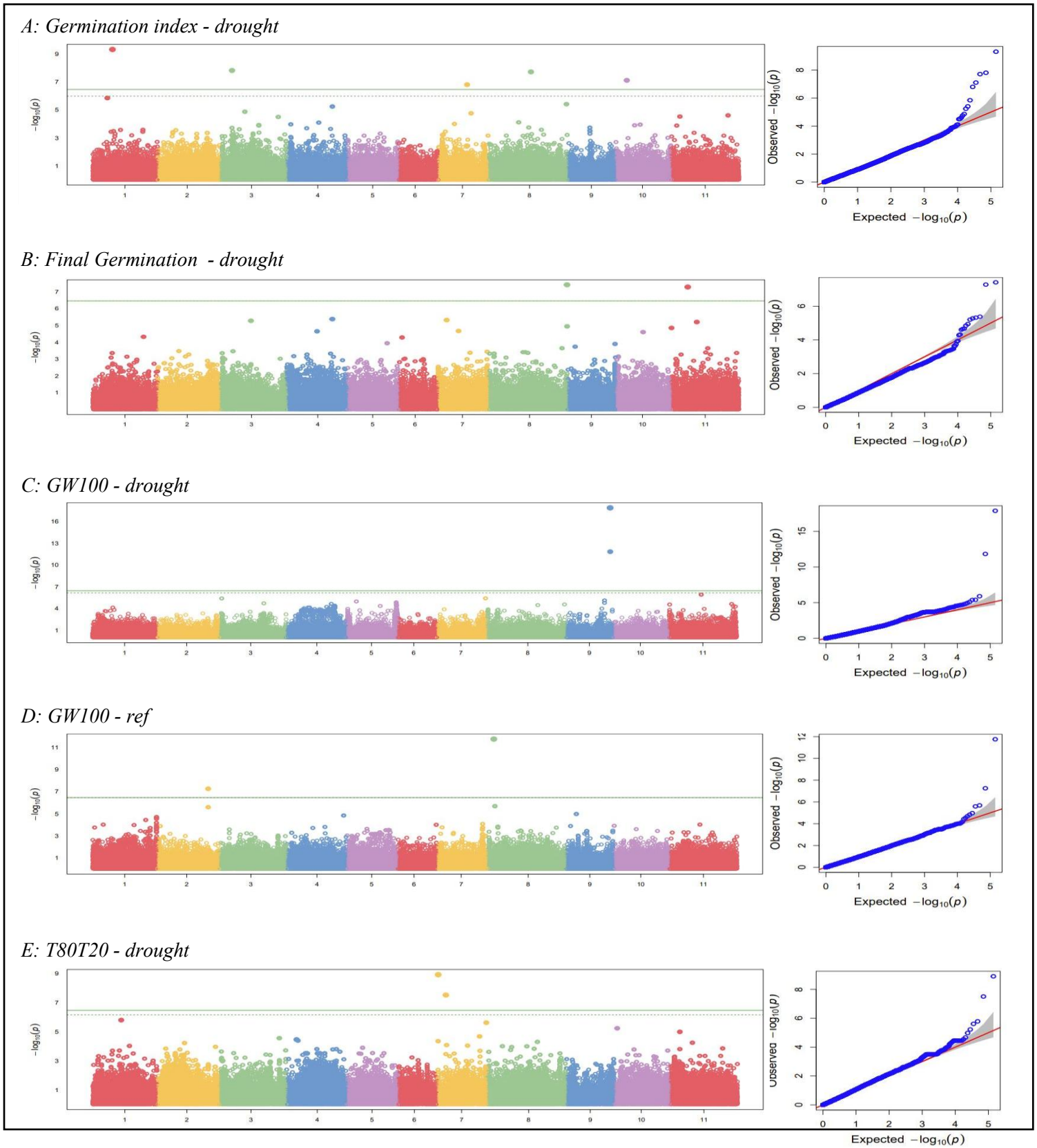


Figure 9: Genome-wide association results across five traits. Manhattan plots show −log10(p) by chromosome from BLINK with PC1–PC5 covariates (n = 170). Horizontal lines indicate genome-wide (solid) and suggestive (dashed) thresholds. Adjacent Q–Q plots show appropriate calibration with trait-specific tail departures at discovered peaks. Lead SNPs correspond to those listed in Table 3.

3.7 LD windows

LD structure around the GWAS index SNPs was broader than initially assumed. Across the 13 focal sites, contiguous high D′ typically extended well beyond 50 kb and decayed at ~100–180 kb from the index, consistent with haplotype blocks of ~200–360 kb. Accordingly, a uniform core window of 300 kb (±150 kb) was adopted for downstream gene prioritization and colocalization (Table 4). This choice aligns with common practice in bean GWAS that define fixed candidate regions in the ±50–100 kb range—e.g., Zia et al. searched 50 kb on each side of significant SNPs and reported panel LD decay starting around ~137 kb, and Moghaddam et al. (2016) centered candidate searches on 100-kb windows—providing precedent for fixed-width windows whose size is tuned to panel-level LD; here, the broader LD observed justifies ±150 kb. The two GW100_drought peaks on Chr09 (35.392 and 35.435 Mb; ~43 kb apart) were treated as a single locus and summarized with a 300-kb merged window centered at 35.414 Mb; the union window (35.263–35.563 Mb) yielded equivalent candidates. All remaining loci (GI_drought on Chr01/03/07/08/10; GR_drought on Chr08/11; GW100_ref on Chr02/08; T80T20_drought on Chr07) were adequately captured by the ±150-kb core windows.

Trait	**Chr**	**Lead SNP (bp)**	**Core window start (bp)**	**Core window end (bp)**
GI_drought	01	15,505,693	15,355,693	15,655,693
GI_drought	03	9,548,597	9,398,597	9,698,597
GI_drought	07	22,968,375	22,818,375	23,118,375
GI_drought	08	33,781,184	33,631,184	33,931,184
GI_drought	10	8,792,003	8,642,003	8,942,003
GR_drought	08	62,503,784	62,353,784	62,653,784
GR_drought	11	13,009,550	12,859,550	13,159,550
GW100_drought	09	35,392,216	35,263,560	35,563,560
GW100_drought	09	35,434,904	35,263,560	35,563,560
GW100_ref	02	40,631,423	40,481,423	40,781,423
GW100_ref	08	5,236,332	5,086,332	5,386,332
T80T20_drought	07	606,868	456,868	756,868
T80T20_drought	07	6,740,141	6,590,141	6,890,141

Table 4: LD-based core windows around GWAS index SNPs.

3.8 Gene annotation

More than 130 different genes were identified across all trait-associated SNPs, distributed within the LD intervals defined for the nine significant loci. The number of potential candidate genes per locus was reduced to a manageable set, by functional inspection of the annotated gene set. This revealed that 36 of the 130 genes were associated with abiotic stress responses, including receptor-like kinases, transcription factors, and enzymes involved in oxidative stress regulation. At least one stress-related candidate was present within the interval of each trait-associated SNP. In addition, several genes with putative roles in seed development or reproductive processes were identified, such as ribose-5-phosphate isomerases, lipid transfer proteins, RNA helicases, and FAR1 transcription factors. The presence of both stress-associated and seed-related genes within the mapped intervals suggests that the loci uncovered in this study capture genetic factors involved not only in adaptation to drought but also in processes directly shaping yield components such as seed weight. (Table 6; Appendix).

3.9 QTL windows

Ten of the thirteen lead SNPs showed colocalization with previously reported QTLs in *Phaseolus vulgaris*. Evidence for seed size and morphology was supported by two loci identified for seed traits: Chr02_40631423_A_G overlapped the seed-weight QTL SW2.1, and Chr08_5236332_G_A aligned with a seed-length region (SL8.1) previously mapped in recombinant inbred populations (Blair et al., 2006; Tar'an et al., 2002). These overlaps are consistent with the roles inferred from the GWAS signals for seed mass–related phenotypes.

Signals linked to drought tolerance and agronomic performance included both chromosome-9 SNPs (Chr09_35392216_T_C and Chr09_35434904_G_T), which fell within a QTL associated with yield under drought and combined stress conditions (Trapp et al., 2015). Additional concordance with broader agronomic or biomass-related regions was observed for Chr03_9548597_G_A (within a multi-trait interval reported from a MAGIC population) and for Chr11_13009550_C_T (within a meta-QTL hotspot for yield components) (Díaz et al., 2020; Pour-Aboughadareh et al., 2022). Chr01_15505693_C_T aligned with intervals reported for morpho-agronomic and seed-quality traits in biparental mapping (Blair et al., 2010).

A domestication-related overlap was detected for Chr07_606868_G_A, which colocalized with a pod shattering resistance region (Shattering7.1**)** encompassing the gene *PvPdh1* (Parker et al., 2020). Pod shattering has been a key trait under selection during domestication, as it directly influences harvestability and seed retention. In the

context of drought response, variation in shattering-related loci may also affect seed release timing, potentially interacting with escape strategies or influencing the onset of germination under dry post-maturity conditions.

For seed quality and germination-related traits, two germination-index SNPs colocalized with QTLs for cooking time (Chr08_33781184_C_T near CT8.2, and Chr10_8792003_G_A near CT10.2) as reported by Cichy et al. (2021). Although cooking time is primarily a post-harvest trait, it is strongly influenced by seed coat permeability and structure—factors that also affect water uptake and seed imbibition during germination. As such, overlaps between cooking time QTLs and germination-associated loci may reflect shared physiological mechanisms related to seed coat composition, hardness, or testa development.

No convincing overlap with published intervals was found for Chr07_6740141_C_T, Chr07_22968375_A_C, or Chr08_62503784_G_A in the literature surveyed.

4. Discussion

The present results indicate that maternal terminal drought generally reduced seed mass and, in many cases, dampened progeny germination performance, but with pronounced genotype-by-environment contingencies. Across 38 accession pairs, seeds produced in the face of terminal drought tended to be lighter, yet, correlations between seed weight and germination metrics (GI, GR, T80T20) were uniformly small and non-significant, and environment × weight interactions were likewise negligible. This pattern suggests that maternal drought alters seed vigor through mechanisms that are not captured by seed mass alone, aligning with reports that stress during grain filling depresses assimilate supply and vigor while producing heterogeneous offspring responses across genotypes (Rosales-Serna et al., 2004; Beebe et al., 2013). Rare cases of equal or superior vigor in seeds produced under drought resemble priming-like "stress memory," consistent with observations in legumes and other species (Damalas et al., 2019; Chen et al., 2021). These inferences are consistent with the accession-level contrasts and multivariate summaries in the dataset (Tables 3; Figs. 2–6).Viewed through the lens of maternal stress memory, these findings fit a framework in which pre- or post-fertilization stress leaves heritable physiological/epigenetic imprints that decouple seed size from vigor traits such as rate and synchrony of germination (Brunel-Muguet et al., 2025).

Several non-exclusive mechanisms plausibly link maternal drought to the observed vigor outcomes. First, drought during maturation can shift hormone and redox balance (ABA/ROS), delaying or desynchronizing germination; the present GWAS intervals include kinases and redox-associated enzymes that could mediate such signaling effects (Beebe et al., 2013). Second, seed-coat physiology likely contributes: maternal stress can harden seed coats, slow imbibition, and broaden the T80T20 window without damaging the embryo, a view supported here by colocalization with cooking-time/seed-quality QTL that reflect coat permeability and texture (Cichy et al., 2021). Third, the outcome depends on developmental timing and remobilization: accessions with "escape/efficient remobilization" strategies are expected to sustain seed filling under terminal stress, whereas lines prioritizing photoprotection without strong late remobilization may still produce inferior seed (Rosales-Serna et al., 2004; Beebe et al., 2013). The strategy-stratified germination curves in this study align with these predictions.

Germination-related phenotypes behaved as polygenic and context-sensitive traits. BLINK-based mapping yielded 13 lead SNPs distributed across GW100_ref, GW100_drought, GI_drought, GR_drought, and T80T20; 9 passed $p < 3.10\times10^{-7}$, with four retained as transparent suggestive signals (Table 3). Gene inventories within these intervals highlighted stress-regulatory modules (receptor-like kinases, redox enzymes, transcription factors) as well as genes tied to seed development, consistent with a dispersed, small-effect architecture commonly reported for seed morphology and

germination vigor in common bean and other crops (Tar'an et al., 2002; Blair et al., 2006; Moghaddam et al., 2016; Wu et al., 2021; Giordani et al., 2022; Shi et al., 2017). The balance of credible biology and modest effect sizes underscores environmental sensitivity and the likelihood that different molecular routes produce similar germination phenotypes under stress. Consistent with the maternal-memory roadmap, overlaying these GWAS intervals with methylome and small-RNA profiles from matched maternal environments could elevate candidates that co-localize with stress-responsive epigenetic features and thereby improve prediction of vigor outcomes (Brunel-Muguet et al., 2025).

Colocalization analyses reinforce biological plausibility and hint at trade-offs. Two GW100_drought peaks on Chr9 overlap drought-yield QTL (Trapp et al., 2015), while a T80T20 signal near 0.61 Mb on Chr7 falls within a domestication region associated with pod shattering resistance around PvPdh1 (Parker et al., 2020). Seed-size loci at Chr2:40.63 Mb and Chr8:5.24 Mb intersect classic QTL for seed weight and length (SW2.1, SL8.1) (Blair et al., 2006; Tar'an et al., 2002). Two GI_drought loci (Chr8, Chr10) align with cooking-time/seed-quality QTL (Cichy et al., 2021), cohering with the seed-coat hypothesis for altered imbibition and spread. Overlaps with broader agronomic/meta-QTL intervals (Blair et al., 2010; Díaz et al., 2020; Pour-Aboughadareh et al., 2022) suggest either pleiotropy or tight linkage among adaptation, seed-quality, and yield components. These intersections provide immediate shortlists for validation while flagging potential linkage drag (e.g., shattering resistance vs. seed traits) that breeders should anticipate.

The strategy-level lens further clarifies phenotypic patterns. Accessions characterized as escape/efficient remobilizers typically buffered both seed filling and progeny vigor, consistent with maintained source–sink continuity during terminal stress. Recovery/resistance lines exhibited mixed outcomes, underscoring that photoprotective capacity alone does not guarantee high-quality seeds in the absence of robust late-stage remobilization. Susceptible lines performed worst, with lower yield and inferior next-generation germination. Stay-green responses were heterogeneous, an outcome that likely reflects diversity in senescence control and source–sink regulation captured in recent surveys (Bengoa Luoni et al., 2019; Beebe et al., 2013; Labastida et al., 2023). Integrating strategy classification with genetic signals thus provides a mechanistic scaffold for interpreting accession-specific departures from average trends (e.g., rare priming-like improvements under drought). Given projected increases in heat waves and terminal droughts, aligning accession-specific stress windows with reproductive timing may help realize the maternal-priming concept for resilient establishment (Brunel-Muguet et al., 2025).

Multivariate analyses separate a "seed filling/size" axis from a "germination dynamics" axis, implying partially independent selection targets. In this panel, phenotypic

PCA attributed ~47% of variance to a weight-dominated PC1 and ~18% to germination timing and synchrony (PC2), while genetic PCA supported clear population structure and justified the use of PC covariates in GWAS. The two-axis view resonates with classical quantitative findings that yield components and seed vigor are related yet not interchangeable (White et al., 1994), and with agronomic studies emphasizing emergence as a practical bottleneck for climate-robust cultivation and northward expansion (Raveneau et al., 2011; Lamichhane et al., 2020). Selection solely on GW100 risks suboptimal establishment, whereas prioritizing early, synchronous germination without maintaining filling could erode yield potential. Consequently, drought-resilient seed quality—the capacity to maintain both seed filling and favorable germination profiles under maternal stress—emerges as a coherent breeding target—co-localized intervals tied to seed-coat properties and stress signaling offer tractable entry points for marker-assisted and genomic selection.

Several limitations shape inference. The number of maternal pairs was modest, and germination assays were performed under controlled laboratory conditions that only approximate field emergence. For GWAS, prespecified zero-coding used to align sparse phenotypes to the full genotype panel may inflate variance and attenuate effect-size interpretability, even with genome-wide covariates. Some signals remain suggestive, and LD-based candidate lists, though compact, are not definitive. These limitations motivate a focused next phase: fine-mapping of lead loci and construction of near-isogenic contrasts; targeted physiology on matched seed lots (coat permeability tests; ABA and ROS profiling) to link candidate pathways with phenotype; reciprocal maternal designs to partition maternal versus zygotic genetic effects; and field emergence trials to calibrate GI/T80T20/GR as predictors of stand establishment. Cross-referencing validated loci against domestication and seed-quality regions (e.g., shattering, cooking time) will help resolve pleiotropy versus linkage and anticipate breeding trade-offs.

In the context of climate change, these results provide an empirical scaffold for developing climate-smart seed lots by identifying stress windows that avoid vigor penalties, validating molecular indicators of favorable imprints, and integrating those markers with GWAS/QTL information to inform breeding and seed-production decisions (Brunel-Muguet et al., 2025).

4.1 Limitations and next steps

Inference is limited by the modest number of maternal pairs (38 accessions) and the small within-accession replication (five seeds per maternal environment), which together reduce statistical power, widen confidence intervals, and make interaction terms (e.g.,

environment × weight, strategy-specific slopes) sensitive to sampling noise. Population structure further lowers the effective sample size for GWAS, such that true small-effect loci may go undetected and some effect estimates may be upwardly biased ("winner's curse") even with PC covariates. External validity is also constrained by context-dependent germination assays performed under controlled conditions; translation to field emergence across soils and temperatures remains to be established. Finally, LD-anchored candidate lists, while compact for a selfing species, still encompass multiple genes, and functional inference is limited by annotation depth and the absence of immediate knock-out resources.

Priority next steps are therefore twofold. Genetic validation should include fine-mapping and near-isogenic contrasts for lead intervals; multi-environment field trials to relate GI, T80T20, and GR to stand establishment; and seed-physiology assays (seed-coat permeability, ABA/ROS quantification) in matched maternal treatments to test mechanistic hypotheses. Cross-referencing validated signals with seed-quality and domestication regions (e.g., cooking time and shattering) will help resolve pleiotropy versus tight linkage and anticipate breeding trade-offs (Cichy et al., 2021; Parker et al., 2020; Pour-Aboughadareh et al., 2022). Epigenetic prospects merit dedicated designs: multigenerational maternal drought and recovery cycles (e.g., F_1–F_3) to test persistence or resetting of "stress memory"; reciprocal crosses to partition maternal versus zygotic contributions; and paired methylome and small-RNA profiling of seeds from drought- versus control-environment plants to identify differentially methylated regions and sRNA signatures associated with vigor (reviews: Crisp et al., 2016; Lämke & Bäurle, 2017; Quadrana & Colot, 2016). Pharmacological demethylation or targeted demethylation in seedlings could provide complementary evidence for causality where feasible. Together, these steps would distinguish heritable epigenetic marks from transient maternal provisioning effects, refine candidate genes within the GWAS windows, and strengthen the link from laboratory phenotyping to field establishment and breeding utility.

5. Acknowledgements

2025, Uppsala

Over the last two years, I studied, moved from Austria to Sweden, and discovered a new culture, one which I learned to cherish. Along the way, there were many adventures, new friendships, and moments that made this period both memorable and rewarding. Within this broader journey, the present thesis project took shape with the support of many people.

I am deeply grateful to my supervisor, Martha Rendón, for her patient, dedicated guidance throughout this project. I also thank Pär Ingvarsson, head of the lab group, for being supportive and welcoming, and for fostering a collaborative research environment. I am thankful to my colleagues at SLU for their help, discussions, and good humour, and to the administrative and technical staff whose steady support kept the work moving.

My heartfelt thanks go to my family for the many efforts that made it possible to get here. To my parents, whose constant encouragement and belief in me have been unwavering—thank you.

6. References

Angioi, S.A., Rau, D., Attene, G., Nanni, L., Bellucci, E., Logozzo, G., Negri, V., Spagnoletti Zeuli, P.L. & Papa, R., 2010. Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. Theoretical and Applied Genetics, 121, pp.829–843. https://doi.org/10.1007/s00122-010-1353-2

Beebe, S.E., et al., 2013. Phenotyping common beans for adaptation to drought. Frontiers in Physiology, 4, 35. https://doi.org/10.3389/fphys.2013.00035

Bellucci, E., Benazzo, A., Xu, C., et al., 2023. Selection and adaptive introgression guided the complex evolutionary history of the European common bean. Nature Communications, 14, 1908. https://doi.org/10.1038/s41467-023-37332-z

Benjamini, Y. & Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), pp.289–300.

Bengoa Luoni, S., et al., 2019. Transcription factors associated with leaf senescence in crops. Plants, 8(10), 411. https://doi.org/10.3390/plants8100411

Blair, M.W., et al., 2006. Quantitative trait loci for seed weight in a Mesoamerican intra-gene pool population of common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics, 112, pp.1141–1150. https://doi.org/10.1007/s00122-006-0211-6

Blair, M.W., et al., 2010. Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (Phaseolus vulgaris L.). Theoretical and Applied Genetics, 121(3), pp.467–484. https://doi.org/10.1007/s00122-010-1328-1

Cakmak, T., et al., 2009. Acceleration of germination and early growth of wheat and bean seedlings grown under various magnetic field and osmotic conditions. Bioelectromagnetics. https://doi.org/10.1002/bem.20537

Chen, Z., et al., 2021. Genome-wide association studies of seed performance traits in response to heat stress in Medicago truncatula uncover MIEL1 as a regulator of seed germination plasticity. Frontiers in Plant Science, 12, 673072. https://doi.org/10.3389/fpls.2021.673072

Cichy, K.A., et al., 2021. QTL mapping of seed quality traits including cooking time in dry bean (Phaseolus vulgaris L.). Frontiers in Plant Science, 12, 670284. https://doi.org/10.3389/fpls.2021.670284

Cokkizgin, A., 2012. Salinity stress in common bean (Phaseolus vulgaris L.) seed germination. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 40(1), p.177. https://doi.org/10.15835/nbha4017493

Damalas, C.A., et al., 2019. Hydro-priming effects on seed germination and field performance of faba bean in spring sowing. Agriculture, 9(9), p.201. https://doi.org/10.3390/agriculture9090201

Díaz, L.M., et al., 2020. Genetic mapping for agronomic traits in a MAGIC population of common bean. BMC Genomics, 21, 737. https://doi.org/10.1186/s12864-020-07213-6

Giordani, W., et al., 2022. Genome-wide association studies dissect the genetic architecture of seed shape and size in common bean. G3: Genes, Genomes, Genetics, 12(4), jkac048. https://doi.org/10.1093/g3journal/jkac048

Huang, M., Liu, X., Zhou, Y., Summers, R.M. & Zhang, Z., 2019. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience, 8(2), giy154. https://doi.org/10.1093/gigascience/giy154

Labastida, D., et al., 2023. Dissecting the genetic basis of drought responses in common bean using natural variation. Frontiers in Plant Science, 14, 1143873. https://doi.org/10.3389/fpls.2023.1143873

Lamichhane, J.R., et al., 2020. Analysis of soybean germination, emergence, and prediction of a possible northward establishment of the crop under climate change. European Journal of Agronomy, 113, 125972. https://doi.org/10.1016/j.eja.2019.125972

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. & Zhang, Z., 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics, 28(18), pp.2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liu, X., Huang, M., Fan, B., Buckler, E.S. & Zhang, Z., 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies (FarmCPU). PLoS Genetics, 12(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Liu, Y., et al., 2020. Effect of germination duration on structural and physicochemical properties of mung bean starch. International Journal of Biological Macromolecules, 154, pp.706–713. https://doi.org/10.1016/j.ijbiomac.2020.03.146

Moghaddam, S.M., et al., 2016. Genome-wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean. The Plant Genome, 9(3), plantgenome2016.02.0012. https://doi.org/10.3835/plantgenome2016.02.0012

Parker, T.A., et al., 2020. Towards the introgression of PvPdh1 for increased resistance to pod shattering in common bean. Dryad [dataset]. https://doi.org/10.25338/B8K03W

Pipan, B. & Meglič, V., 2019. Diversification and genetic structure of the western-to-eastern progression of European *Phaseolus vulgaris* L. germplasm. BMC Plant Biology, 19, 442. https://doi.org/10.1186/s12870-019-2051-0

Pour-Aboughadareh, A., et al., 2022. Meta-QTL analysis for yield components in common bean (Phaseolus vulgaris L.). Plants, 11(1), 128. https://doi.org/10.3390/plants11010128

Raveneau, M.P., et al., 2011. Pea and bean germination and seedling responses to temperature and water potential. Seed Science Research, 21(3), pp.205–213. https://doi.org/10.1017/S0960258511000067

Rosales-Serna, R., et al., 2004. Biomass distribution, maturity acceleration and yield in drought-stressed common bean cultivars. Field Crops Research, 85(2), pp.203–211. https://doi.org/10.1016/S0378-4290(03)00161-8

Segura, V., Vilhjálmsson, B.J., Platt, A., Korte, A., Seren, Ü., Long, Q. & Nordborg, M., 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genetics, 44(7), pp.825–830. https://doi.org/10.1038/ng.2314

Shi, Y., et al., 2017. Genome-wide association study of salt tolerance at the seed germination stage in rice. BMC Plant Biology, 17(1). https://doi.org/10.1186/s12870-017-1044-0

Tar'an, B., et al., 2002. Genetic analysis of seed traits in common bean using a recombinant inbred line population. Crop Science, 42(2), pp.544–552. https://doi.org/10.2135/cropsci2002.5440

Trapp, J.J., et al., 2015. QTL for yield under multiple stress and drought conditions in a dry bean population. Crop Science, 55(4), pp.1596–1607. https://doi.org/10.2135/cropsci2014.09.0601

VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science, 91(11), pp.4414–4423. https://doi.org/10.3168/jds.2007-0980

White, J.W., et al., 1994. Inheritance of seed yield, maturity and seed weight of common bean (Phaseolus vulgaris) under semi-arid rainfed conditions. The Journal of Agricultural Science, 122(2), pp.265–273. https://doi.org/10.1017/S0021859600087451

Wu, L., et al., 2021. Genome-wide association analysis of drought resistance based on seed germination vigor and germination rate at the bud stage in common bean. Agronomy Journal, 113(4), pp.2980–2990. https://doi.org/10.1002/agj2.20683

Zia, B., et al., 2022. Genome-Wide Association Study and Genomic Prediction for Bacterial Wilt Resistance in Common Bean (*Phaseolus vulgaris*) Core Collection. *Frontiers in Genetics*, 13, 853114. https://doi.org/10.3389/fgene.2022.853114

Popular science summary

Beans feed hundreds of millions of people and are a key, affordable source of protein, minerals, and fiber. But in many regions, bean crops rely on rainfall and are hit hard by heat and dry spells. This thesis asks a practical question with big consequences for farmers and breeders: which parts of the bean genome help maintain seed quality and reliable crop establishment when mother plants experience drought—and do lighter seeds from droughted plants necessarily produce weaker seedlings?

To find out, seeds were taken from plants grown either under normal watering or under a short, well-timed drought just before maturity. Seed weight was measured, and simple germination tests tracked how quickly and how completely seeds sprouted. On average, droughted mother plants produced lighter seeds (about 15% lighter), but lighter seeds did not automatically mean poorer germination: across the collection, differences in germination speed and success were mostly explained by the plant's genetic background rather than by seed size alone.

Next, the study scanned the genome to look for "signposts" (DNA variants) associated with seed weight and germination traits. Thirteen genomic regions stood out across the different traits. These regions were then examined more closely, looking ~150,000 DNA letters to either side to list nearby genes and to check whether past studies had linked those areas to useful traits. The shortlists included genes involved in stress responses and seed development, and several regions overlapped with previously known areas for seed size, cooking time (a trait related to the seed coat), and even pod shattering—a domestication trait affecting harvestability. These overlaps make biological sense: for example, changes to the seed coat can influence both how long beans need to cook and how easily seeds take up water to germinate.

What does this mean in practice? First, breeding for drought-resilient beans should not focus only on making seeds heavier. Instead, it should combine seed weight with traits that promote fast, even germination after stress. Second, the genomic regions highlighted here provide concrete starting points for developing DNA markers to help select better lines. Finally, while the lab experiments are informative, field trials are the next step to confirm which genetic signals translate into stronger crop stands under real-world conditions. Put simply: this work narrows the search for drought-smart beans that still establish well—an important piece of making affordable, nutritious food more climate-resilient.

7. Appendix

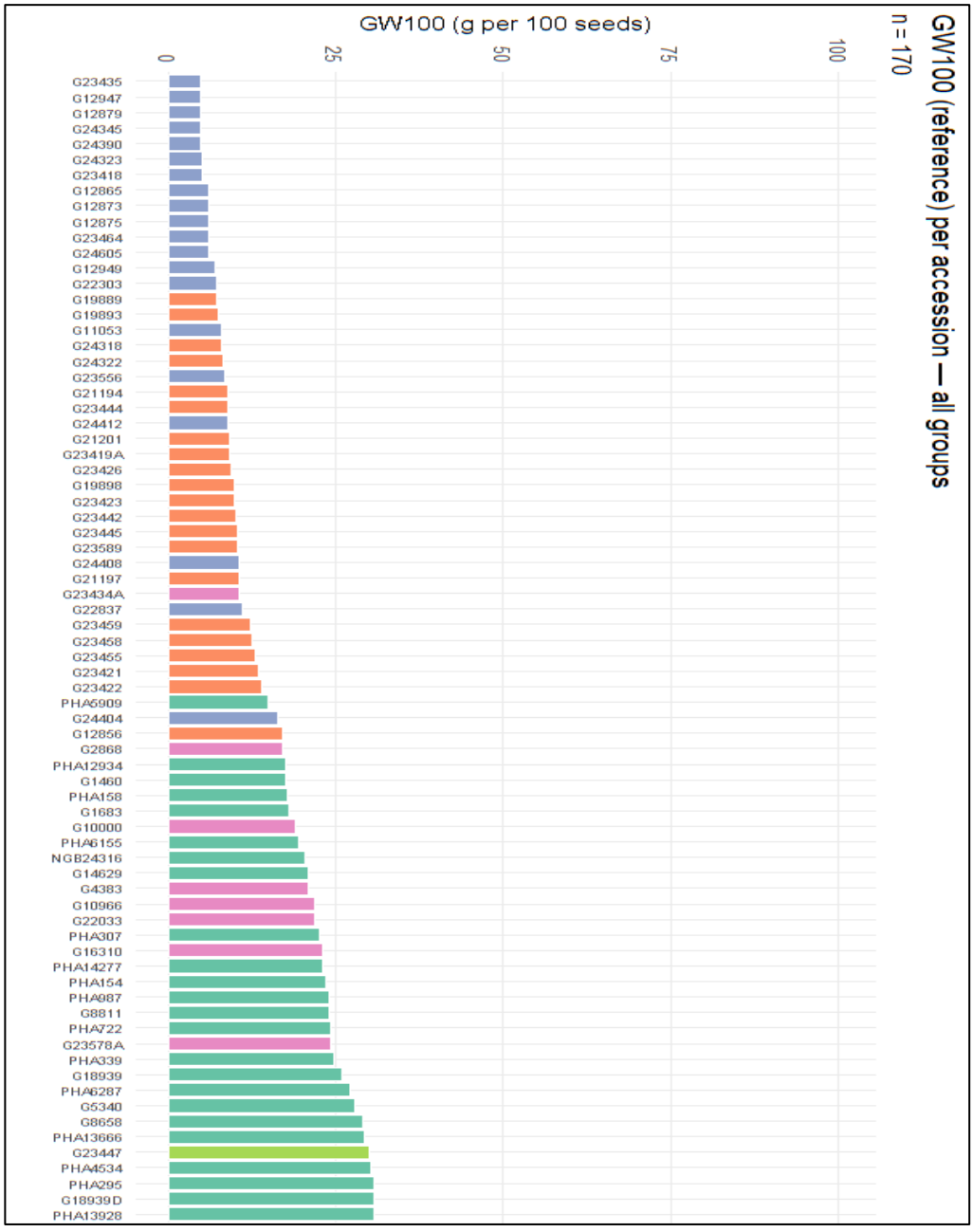
7.1 Accessions

ID	Group	Strategy	ID	Group	ID	Group
G11015	MA	E	G10000	MA	G24345	MW
G1282	EU	R	G10075	EU	G24390	MW
G13094	MA	S	G10093	EU	G24404	MW
G14629	EU	E	G10966	MA	G24408	MW
G23556	MW	S	G11053	MW	G24412	MW
G23578A	MA	R	G12138	A	G24605	MW
G3296	MA	R	G1281	EU	G2868	MA
G4383	MA	S	G12856	AW	G4338	MA
G5340	EU	R	G12865	MW	G4681	MA
G7930	A	R	G12873	MW	G5341	EU
G8658	EU	E	G12875	MW	G7229	A
NGB13468	EU	S	G12879	MW	G8697	A
NGB17826	EU	R	G12947	MW	G8811	EU
NGB18415	EU	SG	G12949	MW	G8920	MA
NGB20124	EU	R	G13177	MA	G900	EU
NGB23857	EU	S	G13614	MA	G9836	A
NGB23858	EU	R	G13948	A	NGB24316	EU
NGB23934	EU	S	G13955	A	PHA109	EU
NGB23936	EU	E	G1460	EU	PHA13112	EU
NGB24038	EU	S	G15914	EU	PHA13181	EU
NGB9300	EU	SG	G16310	MA	PHA13184	EU
PHA1022	EU	R	G1683	EU	PHA13188	EU
PHA1076	EU	E	G16843	A	PHA14277	EU
PHA1077	EU	SG	G18939	EU	PHA143	EU
PHA1086	EU	S	G18939D	EU	PHA154	EU
PHA1137	EU	E	G19889	AW	PHA158	EU
PHA1139	EU	R	G19893	AW	PHA1697	EU
PHA1142	EU	E	G19898	AW	PHA1753	EU
PHA12934	EU	S	G21043	A	PHA1780	EU
PHA13035	EU	R	G21056	A	PHA1887	EU
PHA13099	EU	R	G21069	A	PHA244	EU
PHA13228	EU	S	G21194	AW	PHA2899	EU
PHA13609	EU	S	G21197	AW	PHA295	EU
PHA13666	EU	E	G21201	AW	PHA307	EU

PHA13736	EU	S	G22033	MA	PHA332	EU
PHA13928	EU	E	G22303	MW	PHA339	EU
PHA13960	EU	S	G22837	MW	PHA361	EU
PHA14278	EU	S	G23418	MW	PHA3620	EU
PHA167	EU	R	G23419A	AW	PHA3669	EU
PHA1772	EU	S	G23421	AW	PHA3687	EU
PHA2682	EU	SG	G23422	AW	PHA5877	EU
PHA366	EU	SG	G23423	AW	PHA5909	EU
PHA3673	EU	S	G23426	AW	PHA6287	EU
PHA4008	EU	S	G23434A	MA	PHA722	EU
PHA419	EU	S	G23435	MW	PHA725	EU
PHA4534	EU	E	G23442	AW	PHA7686	EU
PHA4620	EU	E	G23444	AW	PHA841	EU
PHA49	EU	E	G23445	AW	PHA865	EU
PHA5866	EU	S	G23447	A	PHA987	EU
PHA5934	EU	E	G23455	AW	PHA99	EU
PHA5989	EU	R	G23457A	A		
PHA6011	EU	S	G23458	AW		
PHA6066	EU	E	G23459	AW		
PHA6155	EU	SG	G23464	MW		
PHA6254	EU	E	G23589	AW		
PHA6389	EU	S	G23604A	A		
PHA6437	EU	R	G23777	A		
PHA7150	EU	S	G24318	AW		
PHA7309	EU	R	G24322	AW		
PHA7313	EU	R	G24323	MW		

Table 5: List of all 170 accessions in the harmonized panel, with assigned drought-response strategy (Escape, Recovery, Susceptible, Stay-green) and gene-pool (A = Andean domesticated, AW = Andean wild, EU = European, MA = Mesoamerican domesticated, MW = Mesoamerican wild). Strategy annotations are provided where available; remaining entries are to be considered marked NA.

7.2 Weight


GW100 (reference) per accession — all groups
n = 170
GW100 (g per 100 seeds)
0
25
50
75
100
G23435
G12947
G12879
G24345
G24390
G24323
G23418
G12865
G12873
G12875
G23464
G24605
G12949
G22303
G19889
G19893
G11053
G24318
G24322
G23556
G21194
G23444
G24412
G21201
G23419A
G23426
G19898
G23423
G23442
G23445
G23589
G24408
G21197
G23434A
G22837
G23459
G23458
G23455
G23421
G23422
PHA6909
G24404
G12856
G2868
PHA12934
G1460
PHA158
G1683
G10000
PHA6155
NGB24316
G14629
G4383
G10966
G22033
PHA307
G16310
PHA14277
PHA154
PHA987
G8811
PHA722
G23578A
PHA339
G18939
PHA6287
G5340
G8658
PHA13666
G23447
PHA4534
PHA295
G18939D
PHA13928

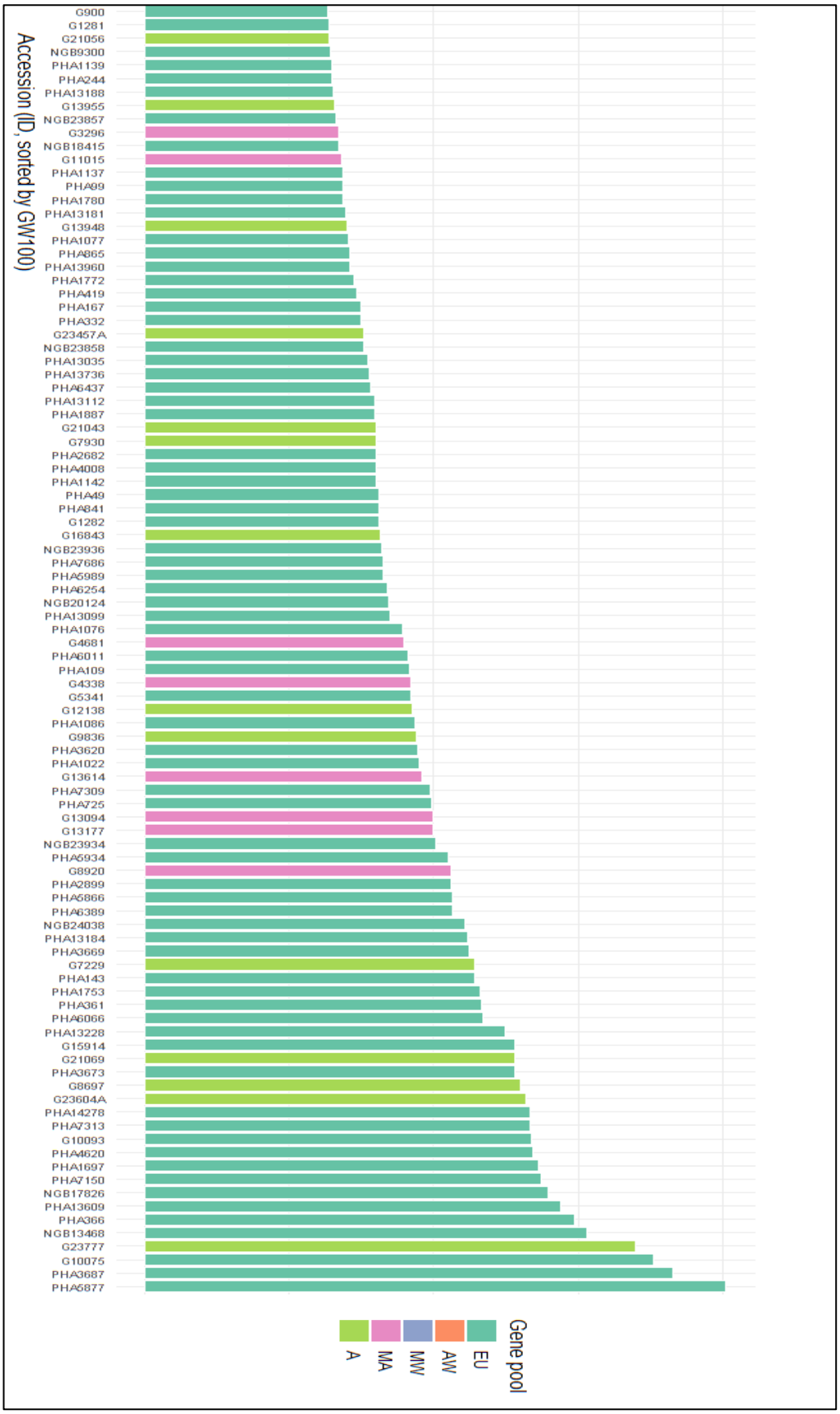
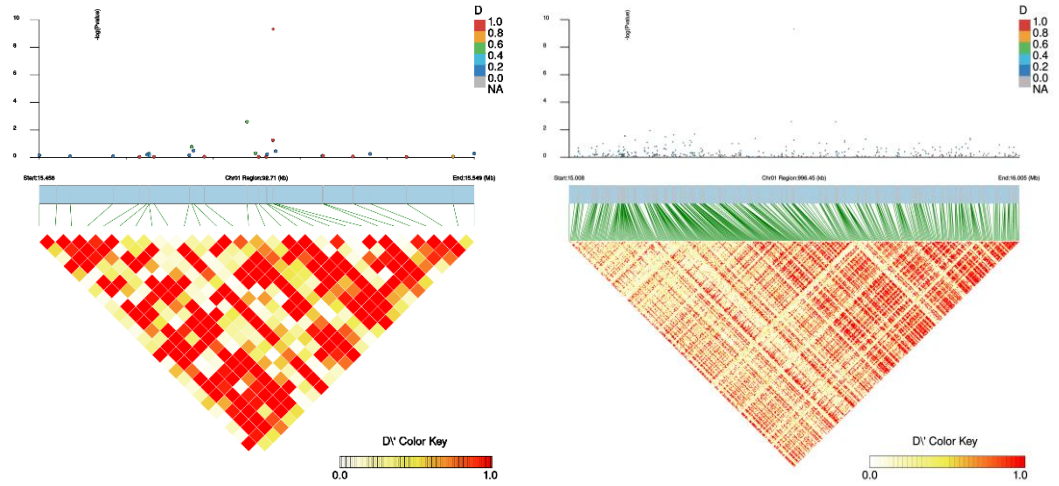


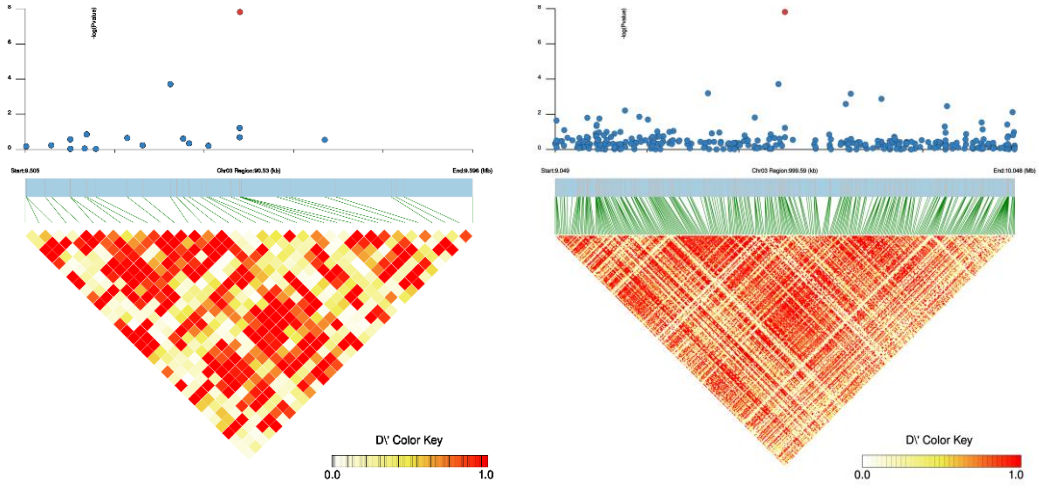
Figure 10: Ordered GW100 by accession, colored by gene pool.

7.3 LD windows

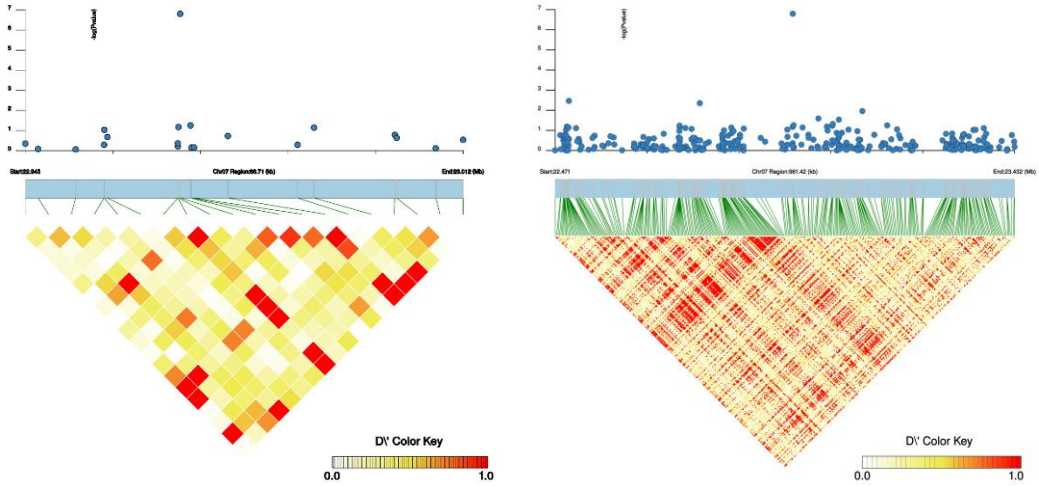
A: Germination Index drought - Chr1 - Pos 15505693



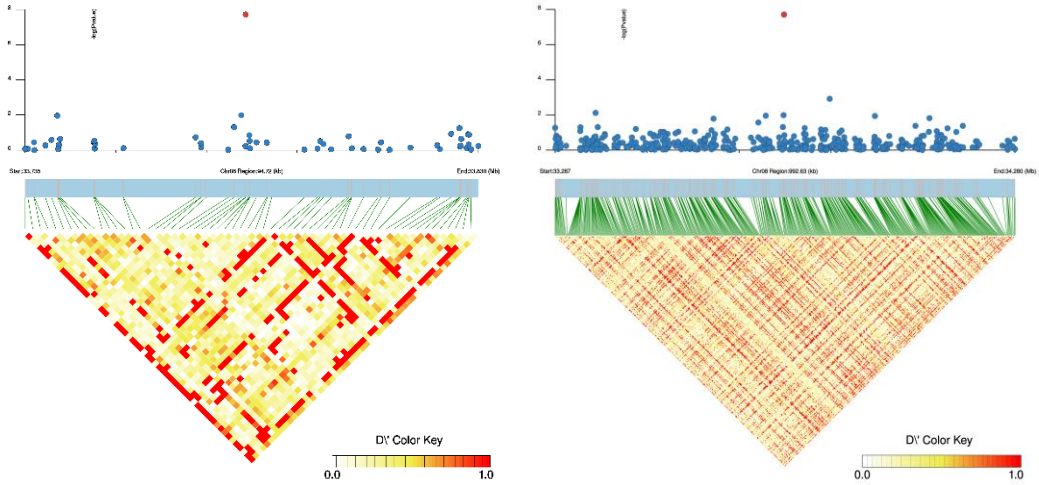
B: Germination Index drought - Chr03 - Pos 9548597



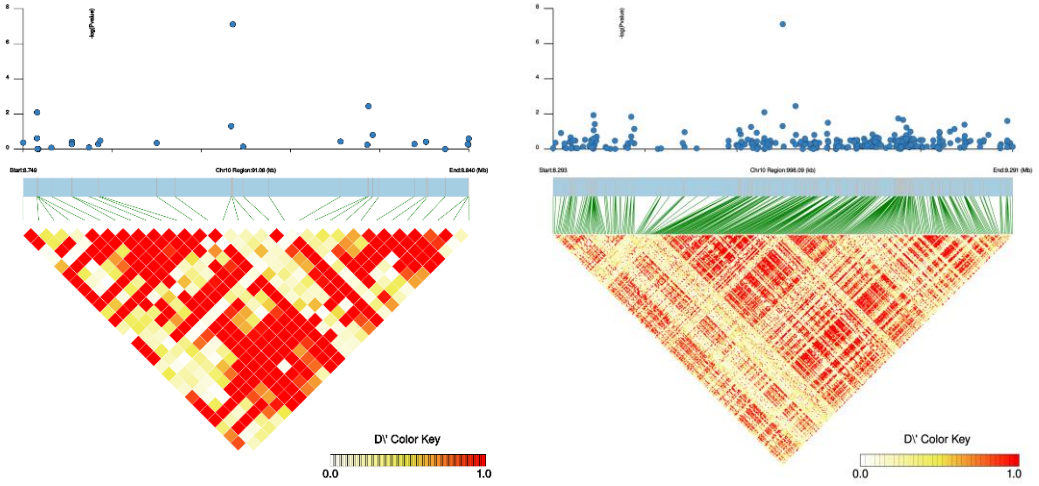
C: Germination Index drought - Chr07 - Pos 22968375



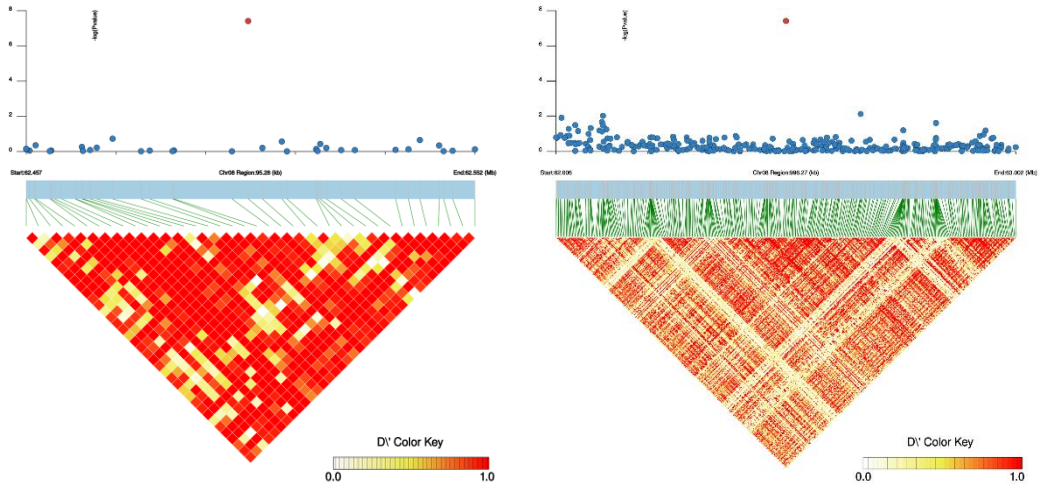
D: Germination Index drought - Chr08 - Pos 33781184



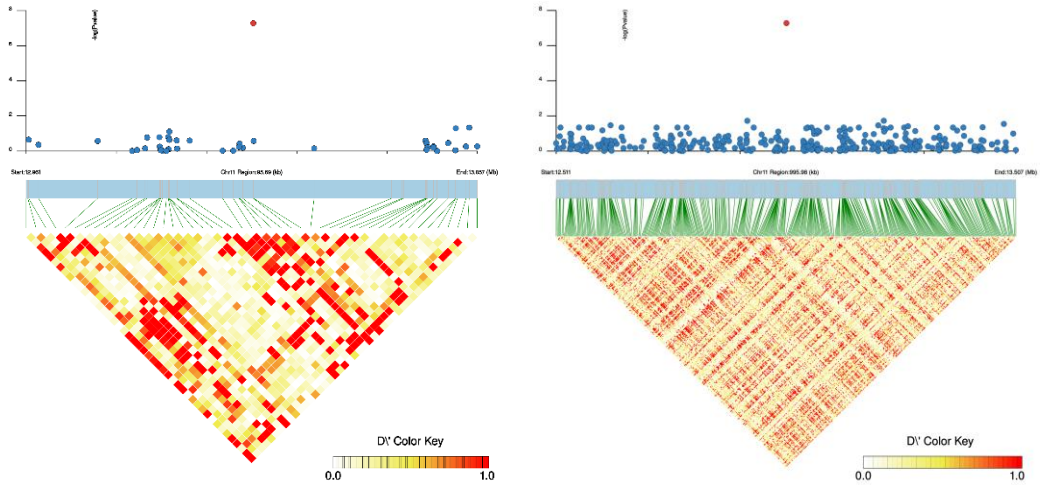
E: Germination Index drought - Chr10 - Pos 8792003



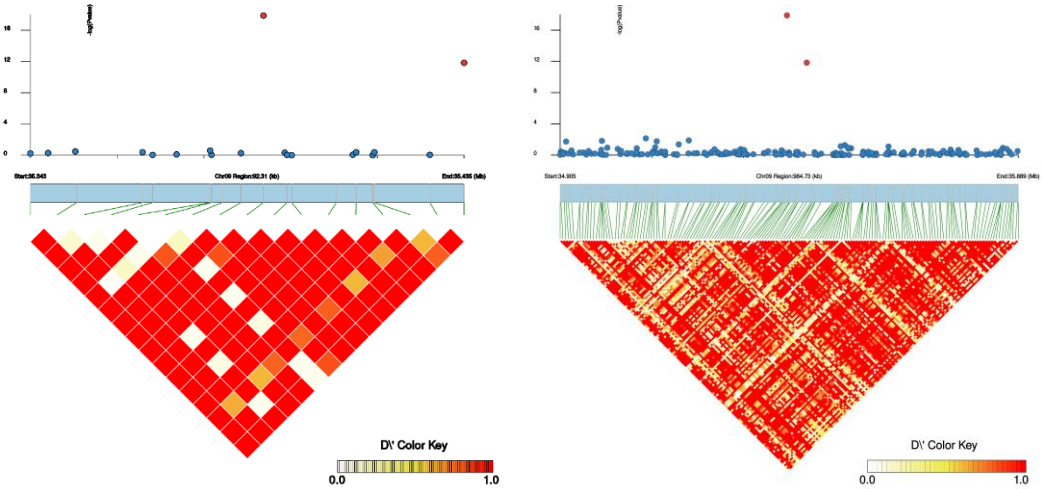
F: Growth Rate drought - Chr08 - Pos 62503784



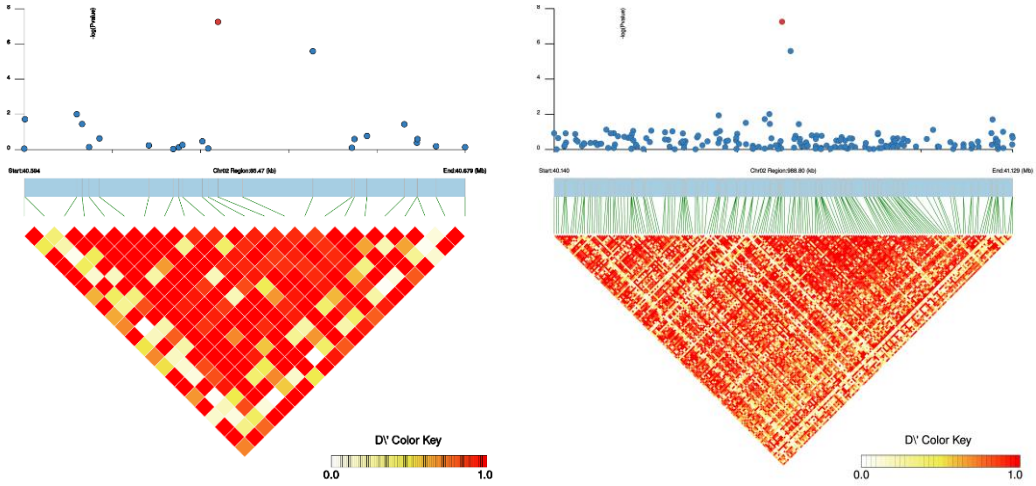
G: Growth Rate drought - Chr11 - Pos 13009550



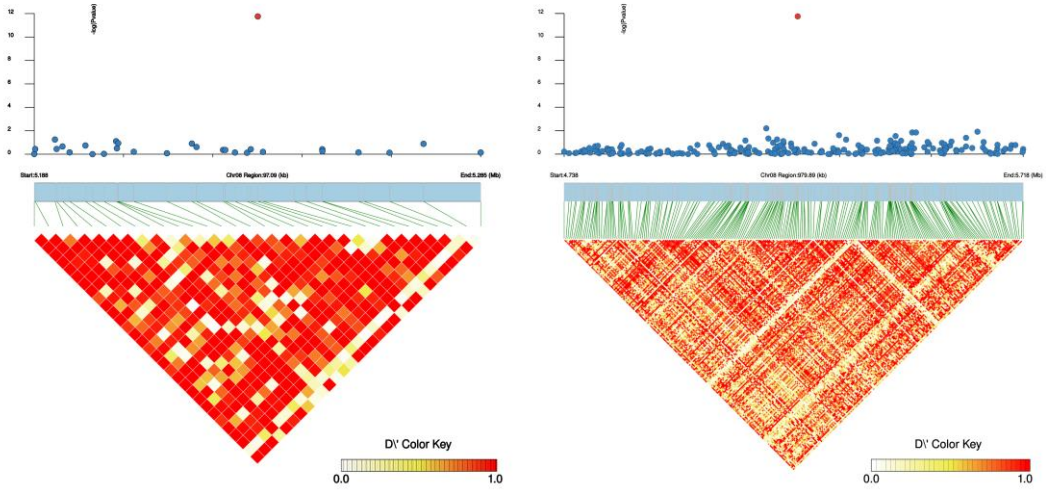
H: GW100 drought - Chr09 - Pos 35392216 – Pos 35434904



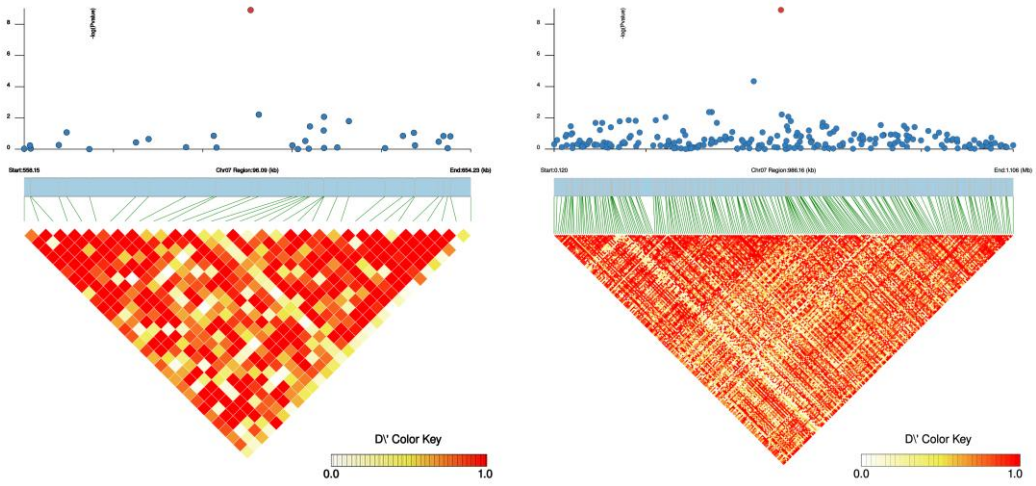
I: GW100 ref - Chr02 - Pos 40631423



J: GW100 ref - Chr08 - Pos 5236332



K: T80T20 drought - Chr07 - Pos 606868



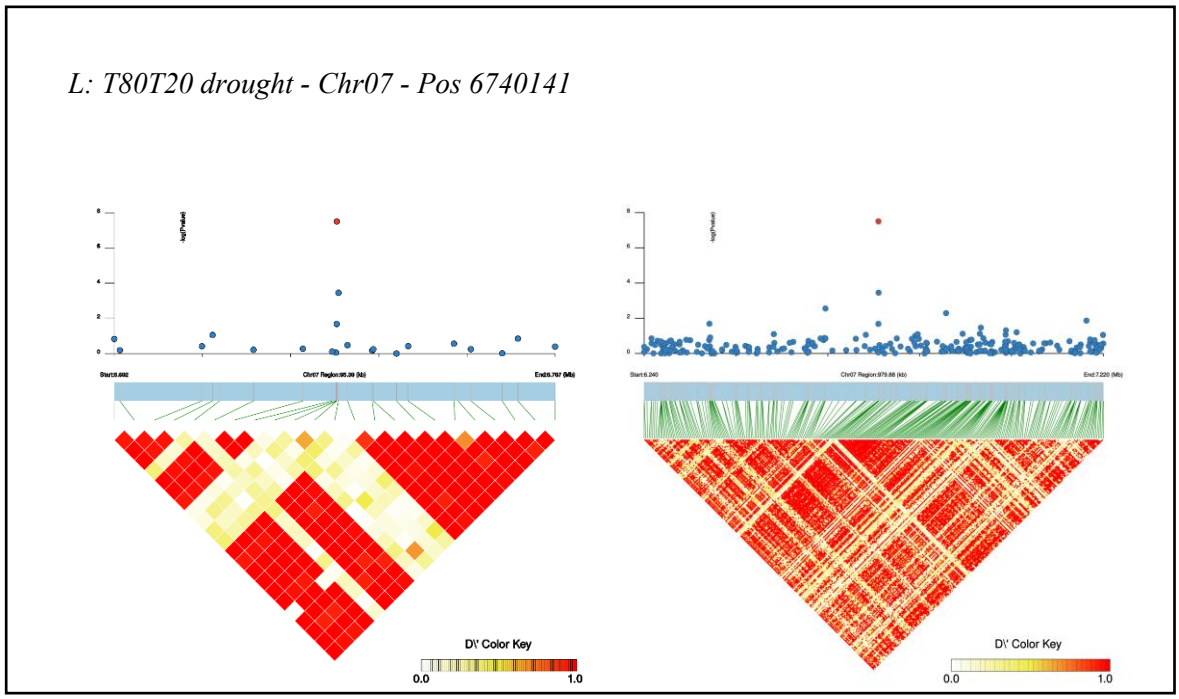


Figure 11: Local and regional LD context for GWAS index SNPs.
For each lead SNP, LDBlockShow heatmaps (D') are aligned with the corresponding Manhattan track. Left panels show ±50 kb windows; right panels show ±500 kb windows. Several loci exhibit extended high-LD segments beyond ±50 kb, motivating the ±150 kb core windows used for gene prioritization.

7.4 Gene annotation

Trait	Phvul ID	Chr:Pos	Arabidopsis orthologs	Function
GR_drought	Phvul.008G285400	Chr08:62,503,784	AT3G07770 (HSP90)	Heat shock protein 89.1 (HSP90 family)
	Phvul.008G284900 / .1	Chr08:62,503,784	AT3G63380 (ACA-type)	Ca^{2+}-transporting ATPase 12, PM-type
GI_drought	Phvul.008G150313	Chr08:33,781,184	AT2G29100 (GLR2.8)	Glutamate receptor 2.8-related
	Phvul.008G136500	Chr08:33,781,184	AT2G44480 (BGLU17)	β-Glucosidase 17
	Phvul.003G111900	Chr03:9,548,597	AT4G16150 (CAMTA5)	Calmodulin-binding transcription activator 5-related
T80T20_drought	Phvul.007G072300.3/.4.v2.1	Chr07:6,740,141	AT3G49060 (PUB32)	U-box domain E3 ligase 32
	Phvul.007G008500 / .1(.v2.1)	Chr07:606,868	AT3G55120 (CHI)	PF02431 → chalcone isomerase
	Phvul.007G008400.1	Chr07:606,868	AT5G05340 (PRX52)	Peroxidase 52
GW100_drought	Phvul.009G236600.1 /.1.v2.1	Chr09:35,392,216	AT4G29950 (TBC1D5)	PF00566 → TBC1D5 (Rab7 GAP)
	Phvul.009G236000 /.1	Chr09:35,392,216	AT1G55770	Invertase/Pectin methylesterase inhibitor family
GW100_ref	Phvul.002G234600 / .1	Chr02:40,631,423	AT2G01290 (RPI2)	Ribose-5-phosphate isomerase 2
	Phvul.002G234900 / .1	Chr02:40,631,423	AT2G42520 (RH37)	DEAD-box RNA helicase 37
	Phvul.008G058500.1	Chr08:5,236,332	AT4G21410 (CRK28)	Cys-rich RLK28-related

Table 6: Annotated candidate genes within GWAS LD windows. For each GWAS signal, the table lists the associated trait, Phaseolus vulgaris gene model (Phvul.), chromosome and genomic position (bp), putative Arabidopsis thaliana ortholog(s), and a concise functional annotation. Genes were collected from LD-based core windows (±150 kb around each index SNP).*

Publishing and archiving

Approved students' theses at SLU can be published online. As a student you own the copyright to your work and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

- https://libanswers.slu.se/en/faq/228318

☒ YES, I, Martin Wohlfahrt, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this.

☐ NO, I/we do not give my/our permission to publish the full text of this work. However, the work will be uploaded for archiving and the metadata and summary will be visible and searchable.