



OPEN Molecular dissection of the genetic architecture of phenology underlying *Lupinus hispanicus* early flowering and adaptation to winter- or spring sowing

Wojciech Bielski^{1,2}, Anna Surma¹, Jolanta Belter¹, Bartosz Kozak³, Michał Książkiewicz¹ & Sandra Rychel-Bielska³✉

Spanish lupin, *Lupinus hispanicus* Boiss. et Reut. is an untapped grain legume species characterized by moderate frost resistance, tolerance to poor soil and waterlogging, high yield stability, and remarkable seed protein content. It has been recognized as a good candidate for domestication to broaden the palette of crop diversity. One of the key characteristics that need to be precisely addressed during domestication is the vernalization responsiveness of flowering, which is advantageous in autumn sowing due to improved frost tolerance, whilst undesired in spring sowing as it delays flowering. Ahead of *L. hispanicus* breeding, in the present work, we aimed to recognize existing phenotypic variability of flowering time and vernalization response and to evaluate the genetic architecture of early and late phenology by DArT-seq genotyping and genome-wide association study (GWAS) in world germplasm collection of the species. Controlled environment phenotyping revealed high variability of flowering time and vernalization responsiveness and significant correlations with population structure. DArT-seq genotyping yielded 23 728 highly polymorphic markers distributed extensively across all 26 chromosomes. GWAS identified a number of markers significantly associated with flowering time with or without pre-sowing vernalization, including those overlapping with the two major quantitative trait loci reported previously for white lupin species. Microsynteny-based analysis of the genetic content of *L. hispanicus* genome regions carrying significantly associated markers highlighted several candidate genes from photoperiodic and vernalization pathways. To summarize, the present study identified germplasm resources for autumn- and spring-sown cultivation of *L. hispanicus* and provided tools for marker-assisted selection towards required flowering phenology.

Keywords *Lupinus hispanicus*, GWAS, Flowering induction, Vernalization, Legumes, Legume crop

Domesticated lupin species are among the most valuable crops due to their natural adaptation to wide range of environments and appreciated grain chemical composition. Their seeds contain high concentrations of essential amino acids and proteins, and they produce stable amounts oils with exceptional nutritional characteristics, including high levels of unsaturated fatty acids and tocopherols. These are important in the human diet and are considered as one of the best alternatives to soybean protein^{1,2}. Additionally, due to the rising global average temperature and changes in precipitation patterns, there is a growing need for plants that can thrive in dry, sandy soils³. This is a typical characteristic of numerous lupin species that differentiates them from many other grain legume crops^{4,5,6,7}. Nowadays, lupins are gaining in popularity as high-protein grain crops in many countries^{8,9,10}.

Lupinus hispanicus Boiss. et Reut., Spanish lupin, is characterized by yield stability, waterlogging, poor soil tolerance, and frost resistance¹¹. Historically, it has been cultivated experimentally^{12,13}. Its high productivity and protein content up to 42%, make it a promising alternative leguminous crop for low-input cultivation^{14,15}. It could also contribute to solving local problems of summer forage scarcity by its use as a self-seeding pasture

¹Department of Gene Structure and Function, Institute of Plant Genetics, Polish Academy of Sciences, Poznan 60-479, Poland. ²Department of Genetics and Plant Breeding, Poznań University of Life Sciences, Dojazd 11, Poznan 60-632, Poland. ³Department of Genetics, Plant Breeding and Seed Production, Wrocław University of Environmental and Life Sciences, Wrocław 50-363, Poland. ✉email: sandra.rychel-bielska@upwr.edu.pl

on the premise that the alkaloid content will be reduced^{16,17}. The *L. hispanicus* is also a valuable genetic resource for introducing beneficial traits into other, closely related lupin species¹⁸. Indeed, the natural crosses between the subspecies *L. hispanicus* subsp. *bicolor* and yellow lupin (*L. luteus*) were observed in Spain, and the first interspecific hybrids were created over 25 years ago¹⁹.

Now, *L. hispanicus* is being considered as a potential new source of plant protein for modern breeding programs¹⁴. Nevertheless, domestication of such an untapped wild plant species will require evaluation of existing germplasm diversity panel for many agronomic traits, which include, among others, the length of period from sowing to flowering and its responsiveness to vernalization, aiming at identification of genotypes with desired phenology. Moreover, the ongoing climate change increases the risk of severe weather events, such as prolonged droughts and heatwave events, that may severely affect regions where lupins and other grain legumes are cultivated^{20,21}. One of the natural strategies to mitigate risk of yield loss in natural lupin populations is drought escape by early phenology, observed in narrow-leaved (*L. angustifolius*), white (*L. albus*) and yellow lupin germplasm and driven by deletions in the promoter region of floral pathway integrator gene, *Flowering locus T*^{22,23,24,25}. It is anticipated that precise adaptation of sowing dates, growing periods and cultivars maturity to local conditions would reduce negative impacts of climate change on crop performance, providing even some yield increase due to the positive CO₂ fertilization effect²⁶. Therefore, one of the biggest challenges for breeding new species is fine-tuning of flowering and maturity dates to given agro-climatic conditions. It is worth noting that the recent climate change already resulted in yield instability of winter crops requiring prolonged chilling period to maintain the process of vernalization and promote transition from vegetative phase to flowering^{27,28,29,30}.

Vernalization responsiveness is one of the key characteristics of many lupin species that needs to be precisely addressed in breeding programs to improve resilience to changing climate³¹. This trait is crucial for many crops, including white lupin, pea, and faba bean, because it provides adaptation to autumn sowing in areas with mild winters with some occurrences of freezing temperatures^{32,33}. In these areas, appropriate fulfillment of moderate or high vernalization requirement not only prevents flowering before the winter but also prevents from frost damage due to a positive correlation between vernalization responsiveness and frost tolerance during winter³⁴. Therefore, the high vernalization requirement of some lupin accessions is one of their most important advantages for autumn or winter sowing in warmer agroclimate, where long juvenile phase is beneficial^{35,36,37}. However, this trait is undesirable for spring sowing, because it may considerably delay flowering when vernalization requirement is not fulfilled due to temperatures above vernalization threshold. Moreover, excessive heat may even erase epigenetic marks and result in late flowering phenotype despite correctly addressed vernalization requirements³⁸.

To summarize, uncovering the mechanism of flowering induction and vernalization response in the new promising candidate crop species is essential for its future breeding. Genetic determinants and their molecular mechanisms underlying observed phenotypic variation should be identified to facilitate reselection of germplasm with desired phenology in the obtained progeny. Several approaches of reduced genome representation sequencing have been successfully implemented in lupins, including massive analysis of cDNA ends (MACE), restriction-site associated DNA sequencing (RAD-seq), genotyping-by-sequencing (GBS) and Diversity Arrays Technology sequencing (DART-seq)^{39,40,41,42,43,44,45}. With recent advances in high-throughput genotyping technologies, genome-wide association studies (GWAS) have become increasingly powerful in detection of sequence polymorphisms associated with various traits, including flowering time and vernalization responsiveness^{4,23,46,47,48}. Finally, the pre-publication release of the chromosome-level genome assembly of a closely related lupin species, *L. luteus*, (GenBank BioProject PRJEB74252, assembly GCA_964019355.1), enabled searching for candidate genes in *L. hispanicus* by shared synteny⁴⁹.

Therefore, in the present study we aimed to characterize the set of *L. hispanicus* lines by reduced genome representation sequencing, to determine their flowering time and vernalization responsiveness, and to associate phenotypic traits with the underlying genetic background by GWAS. The germplasm diversity panel carrying 173 *L. hispanicus* genotypes derived from five seedbanks (located in Poland, Germany, Spain, USA and Australia) was phenotyped in greenhouse for plant phenology under ambient long-day photoperiod with and without pre-sowing vernalization treatment. In parallel, the same set of lines was subjected to DART-seq genotyping, followed by population structure analysis and GWAS exploiting a closely related *L. luteus* genome as a reference (<https://doi.org/10.21203/rs.3.rs-4171664/v1>).

Results

Lupinus hispanicus diversity panel revealed high variability in plant phenology and vernalization requirements

To evaluate *L. hispanicus* phenology, phenotyping of flowering and maturity was performed in greenhouse in years 2022 and 2023 encompassing at least three biological replicates. Detailed information regarding the results, including the number of days from sowing to (a) floral bud emergence, (b) the start of flowering, and (c) pod maturity, as well as growing degree days (GDD) and daylight hours related to the traits mentioned above, can be found in Supplementary Table S1. Below, we provide aggregate statistics for the presented dataset.

Average values per line were calculated, revealing that the number of days from sowing to the appearance of the first floral bud on the non-vernalized plants ranged from 65.0 to 150.0 days in 2022 and from 84.1 to 150.0 days in 2023. To facilitate planned GWAS analysis, lines which remained in good condition until the end of experiment but did not develop any floral bud after 150 days from sowing due to strong vernalization requirements were assigned with values of 150 days for the floral bud emergence, 160 days for the start of flowering and missing data for maturity. The vernalization process accelerated the floral bud emergence in the least responsive accessions by 13.0 days in 2022 and 16.8 days in 2023, whereas in the most responsive genotypes by 94.0 days in 2022 and 89.3 days in 2023. The appearance of the first fully colored petal (start of flowering) on the non-vernalized plants was observed after 79.3 to 160.0 days in 2022 and after 93.6 to 163.0 days in 2023.

Vernalization responsiveness manifested as acceleration of flowering by 15.2 to 93.0 days in 2022 and 16.9 to 90.3 days in 2023. An average number of days from the floral bud emergence to the start of flowering for non-vernalized plants ranged from 7.0 to 21.0 days in 2022 and from 6.7 to 25.0 days in 2023. These values were similar also for vernalized plants, ranging from 6.0 to 17.3 days and from 5.0 to 13.8 days, respectively.

The number of days from sowing to pod maturity on the main stem ranged from 84.3 to 151.0 days in 2022 and from 98.1 to 136.0 days in 2023. The vernalization process reduced these values by 14.5 to 76.3 days in 2022 and by 14.1 to 51.4 days in 2023. It should be noted here, that due to the lack of the floral bud emergence on some non-vernalized plants until the end of experiment, we cannot provide exact values for vernalization-based acceleration of flowering and maturity for the most responsive accessions.

Despite some thermal differences between the two seasons of greenhouse experiments, the earliest genotypes in 2022 were also very early flowering in 2023, and a similar consistency was found for very late genotypes, highly responsive to vernalization. Consistent flowering time patterns were observed for most of the genotypes, with a few exceptions. The most remarkable differences were observed for lines HIS129, HIS142, HIS062 and HIS108, Namely, lines HIS129 and HIS142 in the year 2022 flowered very early, after 79.3 and 83.5 days from sowing (1st percentile), whereas in the next year they flowered over 42 and 40 days later, reaching 64th and 68th percentile in the analyzed set of lines. Oppositely, accessions HIS062 and HIS108 flowered about 36–39 days earlier in 2023, reaching 56th and 67th percentile, than in 2022 (81st percentile).

The average standard deviation of the number of days to bud emergence and start of flowering for non-vernalized plants within accessions reached 3.5 and 4.0 days in 2022 as well as 3.0 and 3.7 days in 2023. For vernalized plants, the corresponding values were 1.8 and 2.4 days in 2022, followed by 1.7 and 2.2 days in 2023. The correlation coefficients for the number of days to the floral bud emergence, start of flowering and maturity between years for non-vernalized plants were 0.906, 0.893 and 0.525, whereas for vernalized plants reached 0.562, 0.596 and 0.530 (Fig. 1). Those correlation coefficients were statistically significant with P-values below 1×10^{-9} . Data on the number of days from sowing to observed phenological phases (floral bud emergence, start of flowering and pod maturity) were provided in Supplementary Table S1. Generalized broad-sense heritability formula, which accounts for both additive and non-additive genetic variances, yielded values between 0.47 and 0.50 for non-vernalized plants and 0.48–0.49 for the vernalized ones. Calculations were conducted on datasets stratified by year to address potential temporal variability, ensuring results were reflective of annual differences. Nevertheless, differences in heritability between years were negligible, as the mean heritability value for all traits was 0.49 in both years.

Phenotypic observations were supplemented with calculation of the cumulative photoperiod hours and the cumulative growing degree days (GDDs) from sowing to reaching particular phenological phase. Thus, for non-vernalized plants in 2022, the cumulative photoperiod hours from sowing to the floral bud emergence, start of flowering and maturity varied from 886 to 2273, from 1115 to 2422, and from 1197 to 2288 h, respectively. For vernalized plants, the corresponding values were from 662 to 1247, from 793 to 1498, and from 891 to 1532 h. This trait was highly correlated (as expected) with the number of days from sowing to the floral bud emergence, start of flowering and maturity in both years (r-value above 0.999).

The cumulative number of GDDs from sowing to the floral bud emergence, start of flowering and maturity for non-vernalized plants varied in 2022 from 1137.7 to 2949.8, from 1411.6 to 3177.0, and from 1500.9 to 2970.7, respectively. For vernalized plants, the corresponding values were from 860.2 to 1562.0, from 1018.0 to 1882.0, and from 1143.7 to 1924.4. Consequently, in 2022, vernalization accelerated the floral bud emergence from 248.8 to 1988.9 GDDs, flowering from 285 to 2002.4 GDDs and maturity from 272.3 to 1644.8 GDDs. In 2023, these values ranged from 339.7 to 1862.2 GDDs for the floral bud emergence, from 350.1 to 1937.6 GDDs for flowering and from 303.0 to 1122.8 GDDs for maturity. This trait was also highly correlated with the number of days from sowing to the floral bud emergence, start of flowering and maturity in both experiments (r-value above 0.999). Data on the number of cumulative growing degree days (GDDs) and photoperiod hours (H) from sowing until reaching studied phenological phases are provided in Supplementary Table S1.

DArT-seq genotyping of *L. hispanicus* diversity panel provided 23,728 high quality markers distributed extensively across all 26 pseudochromosomes

Lupin DArT-seq 1.0 protocol yielded a total number of 125 218 SNP markers and 177 026 presence/absence variants (PAVs, SilicoDArTs), from which ~ 107 000 SNPs and 124 000 PAVs were monomorphic. After filtering by reproducibility rate, missing data, heterozygosity and minor allele frequency (MAF) a total of 5959 SNP and 17 769 PAV markers were retained for population structure analysis and GWAS (Supplementary Table S2). The total number of 13359 PAV and 4352 SNP markers were successfully aligned to *L. luteus* genome assembly, whereas 4410 PAV and 1607 SNP markers remained unlinked. The number of aligned markers per chromosome ranged from 376 on chromosome YL-08 to 965 on chromosome YL-09. The average marker density ranged from 9.56 markers/Mbp (chromosome YL-08) to 25.93 markers/Mbp (chromosome YL-25), with mean value of 18.79 markers/Mbp. Ten chromosomes (YL-25, YL-09, YL-18, YL-26, YL-23, YL-22, YL-12, YL-21, YL-10 and YL-13) reached density more than 20 markers/Mbp. There mean gap between markers was 53.8 kbp, however, 98 gaps were larger than 1 Mbp. Three largest gaps between markers were located on chromosomes YL-02 (4.8 Mbp), YL-07 (4.4 Mbp) and YL-08 (4.4 Mbp) (Supplementary Table S3). After missing data imputation, the mean heterozygosity of markers was 3.97%, ranging from 0.00 to 21.97%, with 17 372 markers showing heterozygosity below 5% and 53 markers above 15%. The heterozygosity of genotypes ranged from 0.31 to 27.48%. Within the analyzed germplasm panel, 137 genotypes had heterozygosity below 5%, whereas 7 above 15% (HIS_122, HIS_024, HIS_007, HIS_123, HIS_116, HIS_064 and HIS_226).

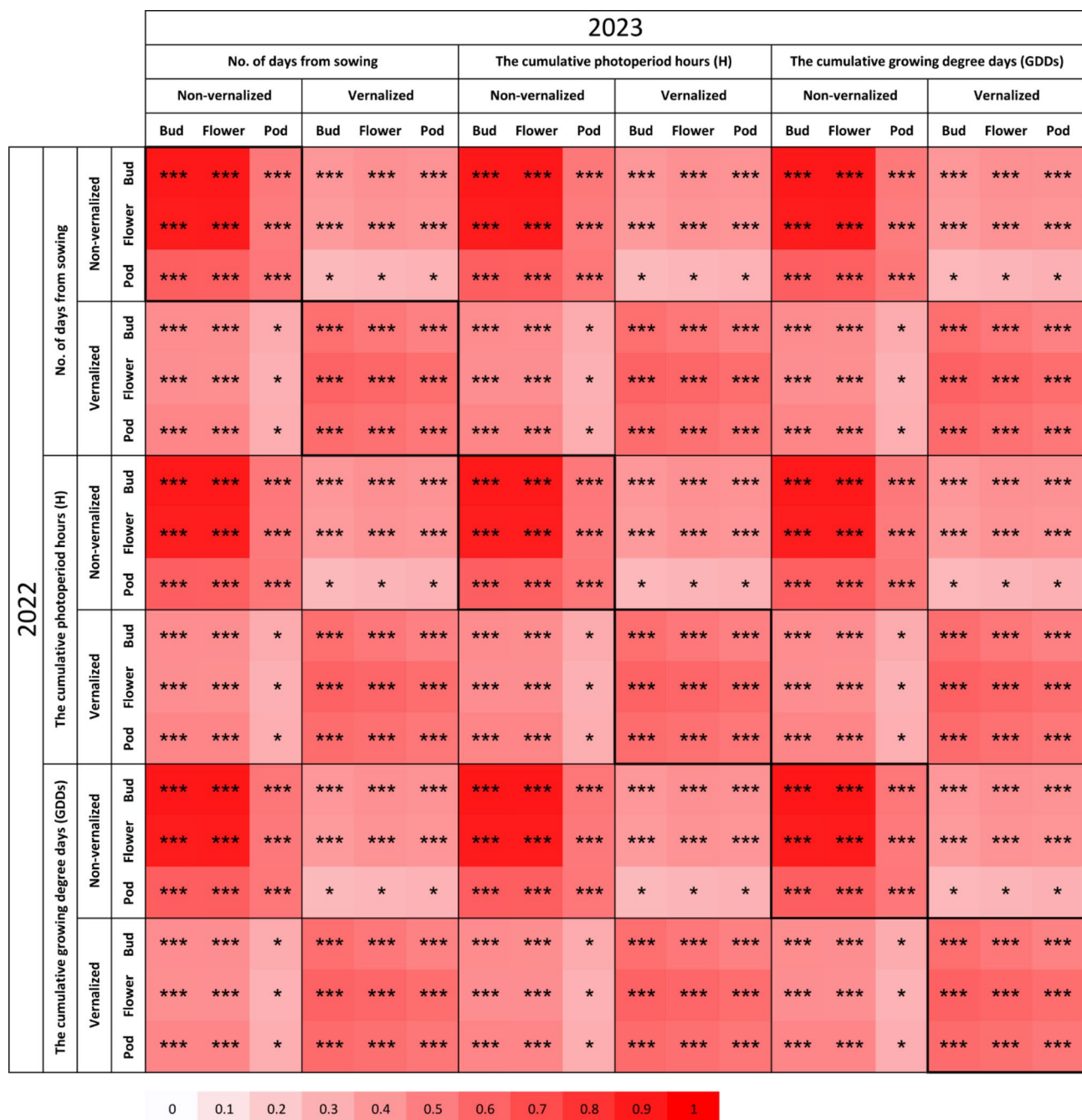


Fig. 1. Correlation heatmap reporting Pearson correlation coefficients between years for each trait. Observations were performed during 2022 and 2023 growing seasons in a greenhouse at the Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland (52°26' N 16°54' E). The bar below the heatmap indicates the color legend of correlation coefficients (from 0 to 1). Asterisk (*) indicates significant correlations in the following scheme: ***, $p < 0.00001$; **, $0.00001 \leq p < 0.0001$; *, $0.0001 \leq p \leq 0.05$.

Plant phenology of *L. hispanicus* diversity panel was significantly correlated with population structure

Analysis of the cross-entropy graph for clustering within the range from K3 to K15 revealed a rapid loss of cross-entropy from K = 3 until a minimum located at K = 6, followed by immediate cross-entropy increase continued until the highest analyzed K-value (Supplementary Figure S1). To validate the optimal K-value selection we performed principal component analysis. Visualization of principal component clusters (PC1 vs. PC2 and PC2 vs. PC3) was provided in Supplementary Figure S2.

K-values within the range of 4–6 were screened for population structure analysis (Supplementary Table S4). Since the first tested K-value (K = 4), grouping started to follow differences in flowering time, especially for non-vernalized plants. As permutation test provides assignment to particular clusters in the continuous range

from 0 to 1, we were able to calculate correlation coefficients between observed plant phenology and cluster assignment values. At $K=6$ value, selected as the most representative according to the cross-entropy criterion, assignment to three clusters (C1, C3 and C5) revealed significant correlations with phenotypic observations (Figs. 2 and 3). Thus, the largest cluster, C1 carrying 77 accessions, was significantly correlated with earliness of non-vernalized plants, demonstrated by r -values -0.54 and -0.55 for the number of days from sowing to the

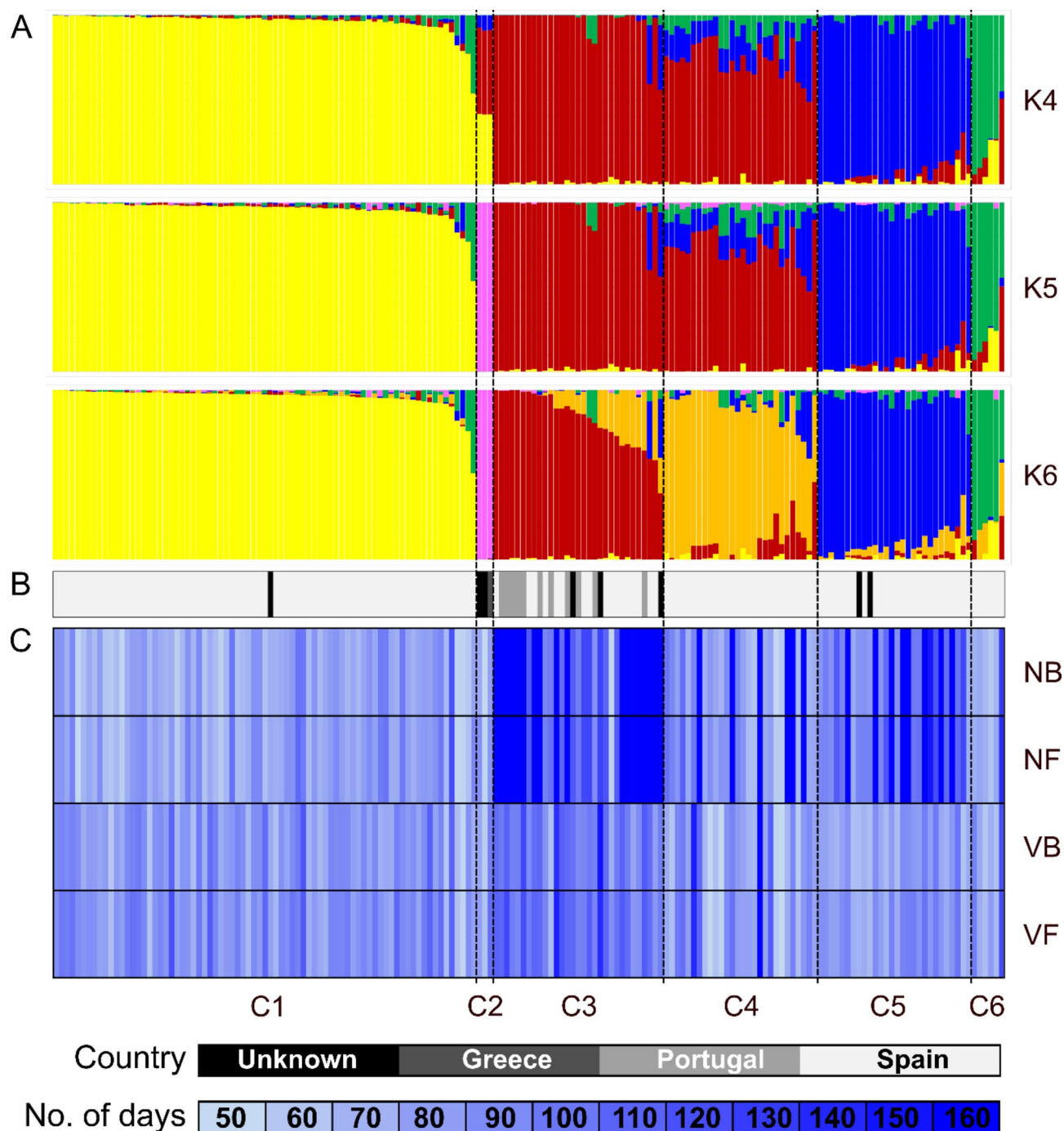


Fig. 2. Visualization of population structure analysis of 173 *Lupinus hispanicus* accessions. The panels show (A) STRUCTURE diagrams under different K-values (K4, K5 and K6), (B) country of origin and (C) mean total number of days from sowing to the floral bud emergence and start of flowering for non-vernalized (NB and NF) and vernalized (VB and VF) plants. In total, 5959 SNP and 17,769 PAV markers were used for population structure analysis. Phenotypic observations were performed during 2022 and 2023 growing seasons in a greenhouse at the Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland (52°26' N 16°54' E). The deeper the shade of blue on the NB, NF, VB and VF panels, the later the phenotype according to the scale provided below the panels.

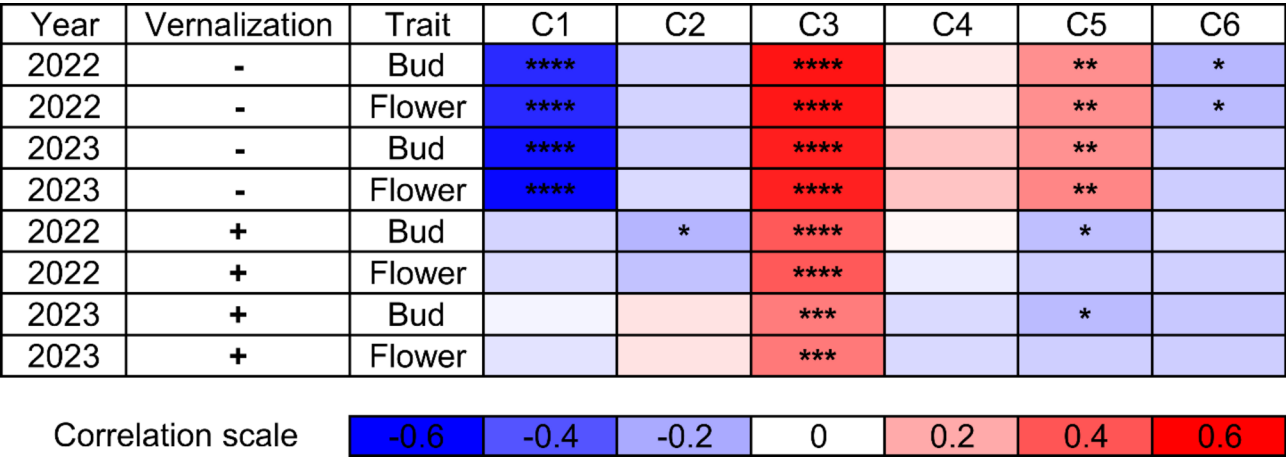


Fig. 3. The Heatmap reporting correlation coefficient values and p-values calculated for assignment to 6 clusters and 8 combinations of trait, year and vernalization. Asterisk (*) indicates significant correlations in the following scheme: ****, $p < 0.00001$; ***, $0.00001 \leq p < 0.0001$; **, $0.0001 \leq p \leq 0.001$; *, $p < 0.05$; “”, non-significant.

floral bud emergence and start of flowering (p -values 1.8×10^{-14} and 2.6×10^{-15}). The second largest cluster, C3 carrying 31 accessions, was significantly correlated with late phenology in both non-vernalized and vernalized variants, manifested by r -values 0.55 and 0.54 (p -values 5.0×10^{-15} and 9.5×10^{-15}) for non-vernalized plants, followed by r -values 0.39 and 0.40 (p -values 6.6×10^{-8} and 5.5×10^{-8}) for vernalized ones. A third largest cluster, C5 carrying 28 accessions, was significantly correlated with late phenotype of non-vernalized plants, however, with remarkably lower parameters (r -values 0.27 and 0.28, p -values 0.0004 and 0.0002) than the two previously mentioned clusters (Fig. 3). Moreover, an *L. hispanicus* accession HIS_232 collected in Greece formed a separate cluster (C2) together with two other accessions with unknown origin (HIS_240 and HIS_241), indicating possible evolutionary diversification by geographic isolation (Fig. 2).

GWAS highlighted several loci significantly associated with *L. hispanicus* phenology and vernalization responsiveness

Based on the PCA and cross-entropy analysis, $K=6$ was selected as the representative number of clusters for population structure in GWAS (Supplementary Figures S1 and S2). Two years (2022 and 2023) and two experimental variants, non-vernalized (N) and vernalized (V) were analyzed. Analyzed trait included the number of days from sowing to the floral bud emergence (NB_2022, NB_2023, VB_2022 and VB_2023), start of flowering (NF_2022, NF_2023, VF_2022 and VF_2023) and pod maturity (NP_2022, NP_2023, VP_2022 and VP_2023). Moreover, the number of days was recalculated as the cumulative number of growing degree days until the floral bud emergence (traits NB_GDD_2022, NB_GDD_2023, VB_GDD_2022 and VB_GDD_2023), start of flowering (NF_GDD_2022, NF_GDD_2023, VF_GDD_2022 and VF_GDD_2023) and pod maturity (NP_GDD_2022, NP_GDD_2023, VP_GDD_2022 and VP_GDD_2023). The same recalculation was done for the total photoperiod hours, providing another set of traits (NB_H_2022, NB_H_2023, VB_H_2022, VB_H_2023, NF_H_2022, NF_H_2023, VF_H_2022, VF_H_2023, NP_H_2022, NP_H_2023, VP_H_2022 and VP_H_2023, respectively). In total 36 variables were analyzed in GWAS. As 25.4% markers did not match any of the *L. luteus* scaffolds, we compared the results obtained with two algorithms, FarmCPU and BLINK, as the first one accounts the genomic position of markers whereas the second one does not analyze it.

The total number of 78 markers revealed at least one marker-trait association with FDR-corrected P -value below the 0.05 threshold, including 47 markers found by BLINK, 40 by FarmCPU and 9 markers significant by both methods (Supplementary Table S5). As many as thirty three markers revealed significant associations with three variables by one of the algorithms, however, usually it was just association with the number of days for one phenological trait and with cumulative GDDs and photoperiod hours calculated for this trait. Therefore, to focus on the more universal markers we looked for the shortlisted thirteen markers showing significant BLINK or FarmCPU associations with at least four variables (Table 1). This set included twelve PAV markers and one SNP marker. Moreover, six markers from this list were identified as significantly associated by both algorithms. Minor allele frequency (MAF) ranged from 6.3% to 9.9% (markers M065429 and M055339) to 47.3%, 47.6% and 48.2% (markers N114410, M089187 and M092420). Marker significantly associated with plant phenology without vernalization were shown not significant after vernalization and vice versa. One marker, M074938, was significantly associated with plant phenology in both years for non-vernalized plants. Employing the Best Linear Unbiased Estimates (BLUEs) method within GWAS, which integrates phenotypic data across multiple years, has confirmed the significant association of four previously selected markers (M065429, M079035, M055339, and M102006), as well as revealed an additional one: M076759. Interestingly, the M076759 marker exhibits a remarkable similarity to M074938, both in terms of position on *L. luteus* chromosome YL-12 (4791451–4791519 bp vs. 4791386–4791452 bp) and allele segregation profile within our *L. hispanicus* diversity panel. The results of BLUE-based GWAS are provided in Supplementary Table S6.

Marker	L. luteus chromosome	Position	Minor allele frequency (%)	No. of BLINK associations	No of Farm CPU associations
M104942	YL-16	23 961 978	24.4	9	7
M079035	YL-23	17 472 037	25.6	5	6
M102006	YL-22	28,557,468	33.1	5	6
M074938	YL-12	4,791,386	18.7	8	2
M055339	YL-03	35,159,449	9.9	3	6
M079300	YL-26	20,608,686	24.1	6	3
M065429	YL-25	17,598,995	6.3	0	6
M089187	-	-	47.6	6	0
N114410	YL-16	7,673,803	47.3	0	6
M076759	YL-12	4,791,451	19.1	6	-
M103926	-	-	26.2	5	0
M075705	YL-04	8,662,920	23.8	0	4
M092420	YL-19	29,238,434	48.2	4	0
M103410	-	-	26.8	4	0

Table 1. Markers that revealed significant associations with at least four variables in *Lupinus hispanicus* diversity panel.

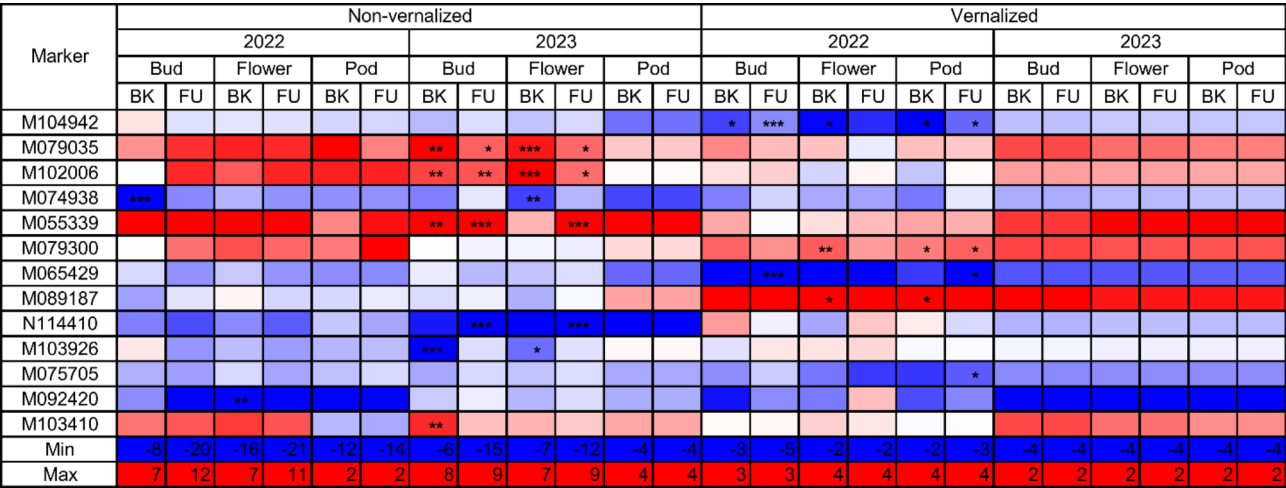


Fig. 4. Marker effects for 13 markers significantly associated with at least four variables. Observed traits include the number of days from sowing to the floral bud emergence, start of flowering and pod maturity for non-vernalized and vernalized plants. Phenotypic observations were performed during 2022 and 2023 growing seasons in a greenhouse at the Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland (52°26' N 16°54' E). Color scale is presented from blue (minimum value) through white (zero) to red (maximum value). BK stands for BLINK, whereas FU for FarmCPU.

The effects of significantly associated markers for non-vernalized plants and the number of days to the floral bud emergence ranged from – 8.0 to 12.0 days in 2022 and from – 15.1 to 11.8 days in 2023, whereas for the start of flowering ranged from – 16.5 to 12.0 days in 2022 and from – 12.4 to 9.5 days in 2023. Corresponding values for GDDs until the floral bud emergence reached from – 181.3 to 283.4 in 2022 and from – 247.4 to 231.4 in 2023, whereas for GDDs until flowering ranged from – 392.6 to 219.0 and from – 315.2 to 204.6, respectively. Similarly, these effects for the total photoperiod hours until the floral bud emergence ranged from – 236.0 to 154.4 h in 2022 and from – 186.8 to 174.4 h in 2023, whereas until flowering reached values from – 348.9 - in 2022 to 146.2 h in 2023.

For vernalized plants, significantly associated markers were identified only for the year 2022, but also for the pod maturity. The effects of those markers ranged from – 4.6 to 2.0 for the number of days from sowing until the floral bud emergence, from – 2.9 to 4.3 days for flowering, and from – 3.3 to 4.1 days for the pod maturity. Effects for GDDs reached from – 81.3 to 47.9 for the floral bud emergence, from – 48.1 to 87.5 for flowering and from – 58.2 to 78.9 for the pod maturity, whereas effects for the total photoperiod hours ranged from – 72.3 to 30.7, from – 46.7 to 68.6 and from – 57.2 to 67.8, respectively. Visualization of effects for the number of days from sowing to the floral bud emergence, start of flowering and pod maturity was provided on Fig. 4. Circular visualizations of Manhattan plots for GWAS were provided in Supplementary Figure S3.

Besides the effects reported by GWAS accounting for population structure, we analyzed the distribution of GDDs recorded for both alleles and heterozygotes of selected 13 significant markers in the germplasm diversity panel (Figs. 5 and 6). The direction of GDD effects for an alternative allele was coherent between years and traits for all markers in non-vernalized plants and for all markers except M074938 and M089187 in vernalized plants. Heterozygotes revealed phenotypes intermediating between those observed for the two opposite homozygotes, indicating incomplete dominance. It should be noted direction of allelic effects reported by BLINK and FarmCPU programs (Fig. 4) were coherent with GDD effects in germplasm diversity panel for all significant marker-trait associations (Figs. 5 and 6). However, for some non-significant marker-trait associations in non-vernalized plants, they were opposite (i.e. markers N114410, M103926, M079035 and M055339).

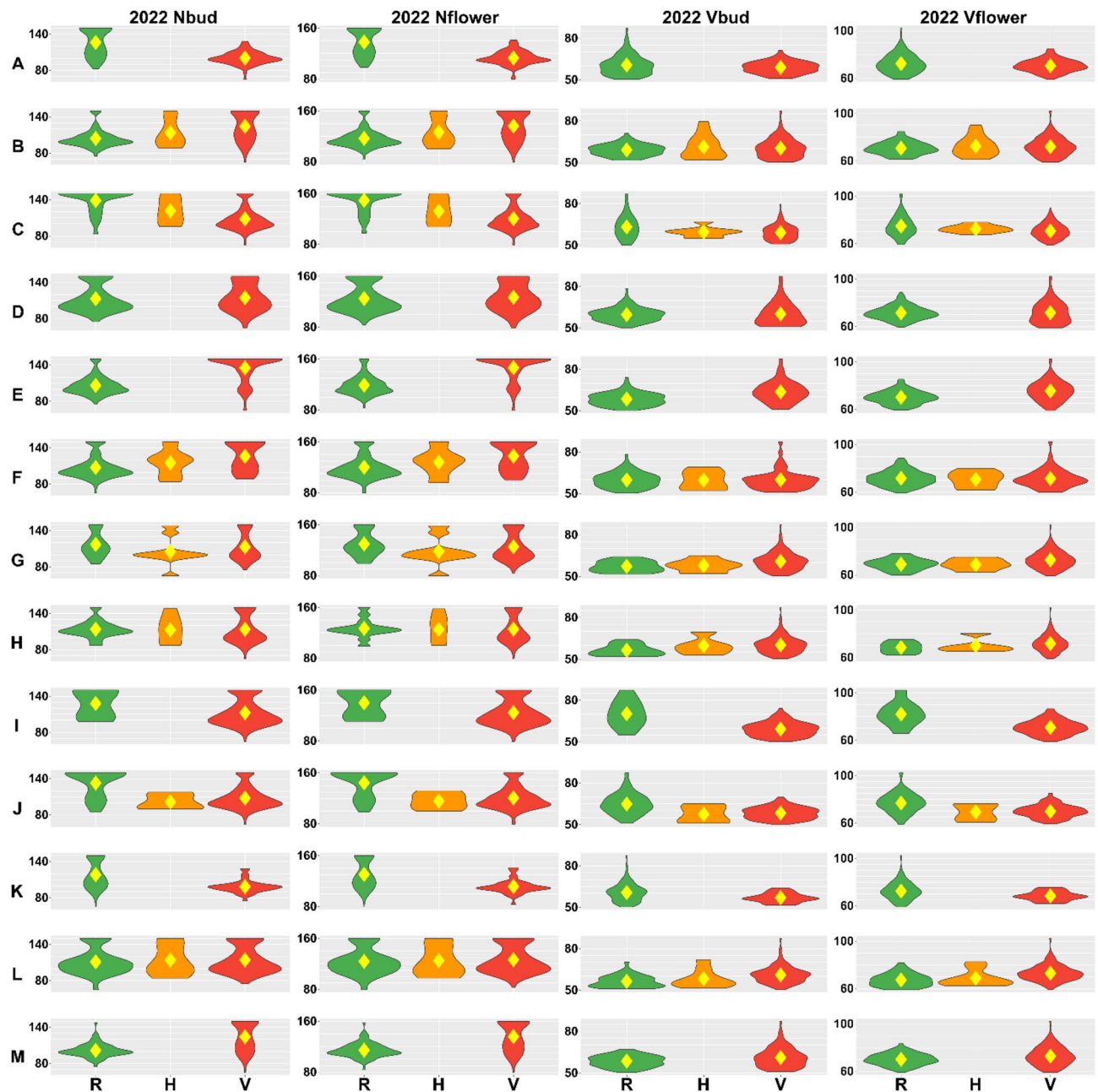


Fig. 5. Allelic effects on the number of growing degree days (GDDs) for thirteen DArT-seq markers, M092420 (A), N114410 (B), M074938 (C), M103926 (D), M103410 (E), M102006 (F), M079035 (G), M055339 (H), M065429 (I), M075705 (J), M104942 (K), M079300 (L), and M089187 (M) tagging novel QTL regions. R stands for the reference allele (0), V for an alternative allele (2), whereas H for a heterozygote (1). Observed traits include the number of days from sowing to the floral bud emergence (bud), start of flowering (flower) and pod maturity (pod) for non-vernalized (Nbud, Nflower, Npod) and vernalized plants (Vbud, Vflower, Vpod). Phenotypic observations were performed during 2022 growing season in a greenhouse at the Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland. Diamonds indicate mean values.

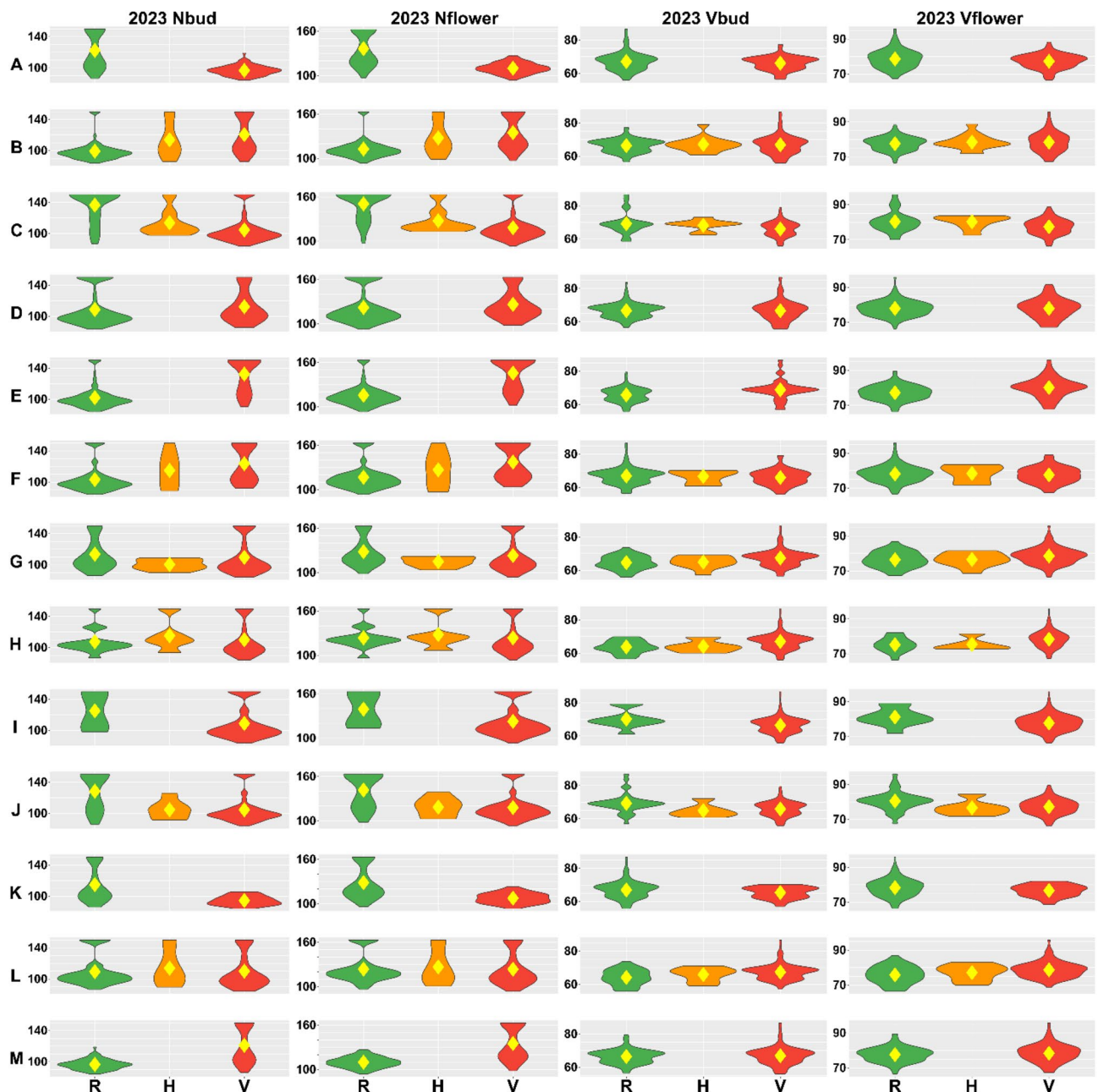


Fig. 6. Allelic effects on the number of growing degree days (GDDs) for thirteen DArT-seq markers, M092420 (A), N114410 (B), M074938 (C), M103926 (D), M103410 (E), M102006 (F), M079035 (G), M055339 (H), M065429 (I), M075705 (J), M104942 (K), M079300 (L), and M089187 (M) tagging novel QTL regions. R stands for the reference allele (0), V for an alternative allele (2), whereas H for a heterozygote (1). Observed traits include the number of days from sowing to the floral bud emergence (bud), start of flowering (flower) and pod maturity (pod) for non-vernalized (Nbud, Nflower, Npod) and vernalized plants (Vbud, Vflower, Vpod). Phenotypic observations were performed during 2023 growing season in a greenhouse at the Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland. Diamonds indicate mean values.

As the majority of *L. hispanicus* DArT-seq markers were localized on the pseudochromosomes of the closely related species *L. luteus*, we were able to analyze linkage disequilibrium (LD) decay around marker loci. From the 14 markers significantly associated with at least four variables in GWAS (including an additional one from BLUE analysis), eight were located in regions well saturated with other DArT-seq loci (≥ 20 markers per 1 Mbp) and as such were subjected to LD analysis (Fig. 7).

Markers N114410 from chromosome YL-16, M092420 from YL-19 and M102006 from YL-22 were localized in the middle of the regions exhibiting high LD values on large (> 1 Mbp) chromosomal segments. Markers M074938 from chromosome YL-12, M065429 from YL-25 and M079300 from YL-26 showed rapid LD decay with chromosomal distance, whereas a marker M075705 from chromosome YL-04 revealed intermediate values

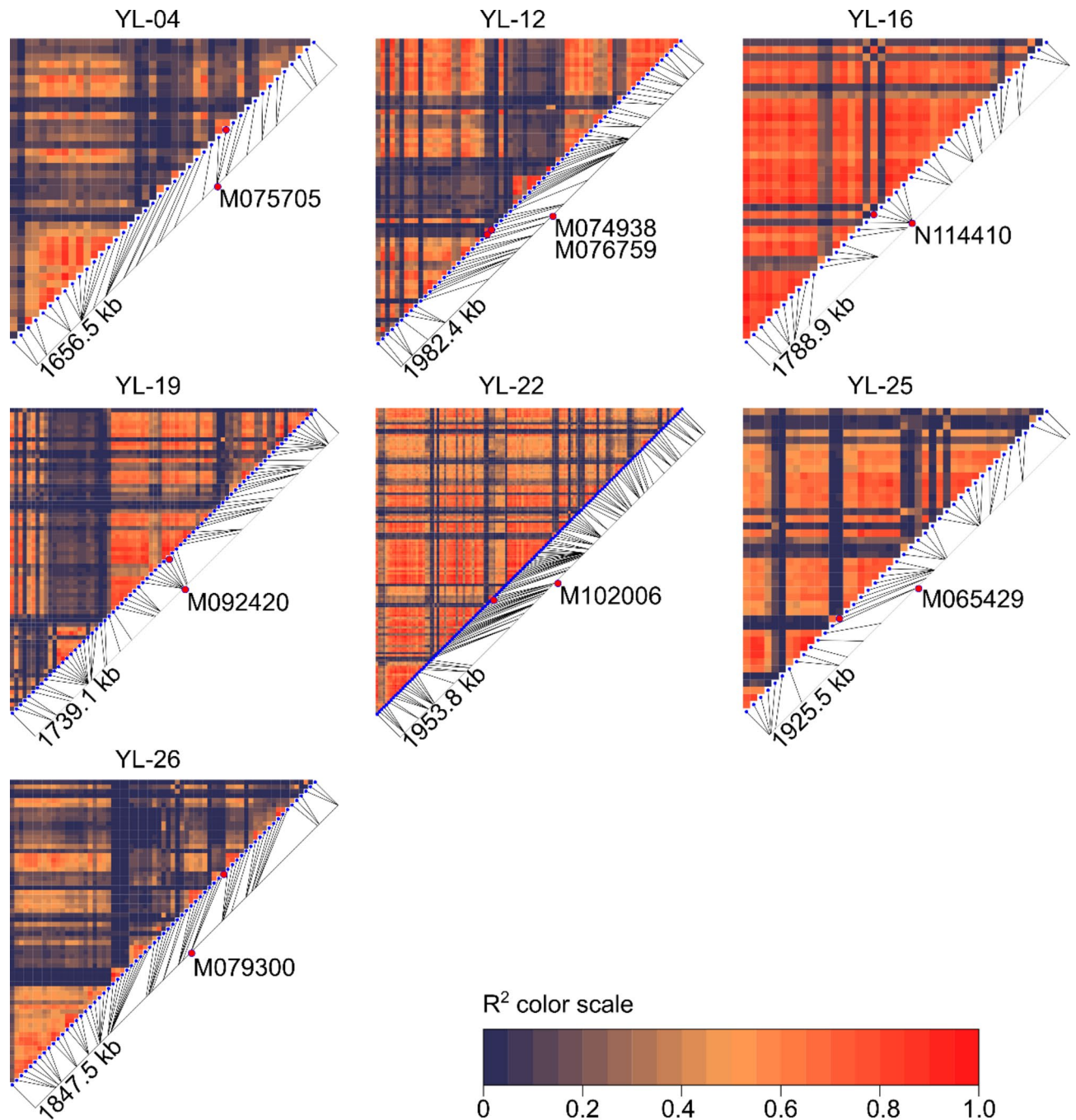


Fig. 7. Linkage disequilibrium (LD) plots for *L. luteus* genomic regions carrying DArT-seq markers significantly associated with *L. hispanicus* plant phenology. The pairwise r^2 values between markers are shown. Red indicates high measures of LD, while deep violet indicates low LD. Lengths of analyzed segments is provided on the marker bars.

of LD. However, rapid LD decay around some marker loci may result from using a proxy *L. luteus* genome for sequence alignments and the possibility of read mapping into incorrect locations or the presence of recombination hotspots causing break of microsynteny between these two species.

Candidate genes identified by shared microsynteny

Due to the lack of sequenced *L. hispanicus* genome and raw annotation of pre-publication *L. luteus* genome release (GCA_964019355.1), we exploited shared microsynteny between *L. luteus* scaffolds carrying significant *L. hispanicus* markers and the reference *L. albus* pangenome assembly to identify candidate genes for markers showing significant BLINK or FarmCPU associations with at least four variables^{50,51}. Alignment of DArT-seq sequences revealed gene-specific hits for all 14 selected markers in the *L. albus* genome and for all markers except

M104942 in the *L. luteus* genome. According to the annotations provided in the *L. albus* genome assembly, those genes putatively encode such proteins as Zf-FLZ domain-containing protein, pre-mRNA polyadenylation factor Fip1, pectinesterase/pectinesterase inhibitor 41, ribosomal protein L34Ae, cellulose synthase (UDP-forming), transcription initiation factor IIB, Pumilio RNA binding domain-containing protein, cytochrome P450, DNA helicase, serine/threonine kinase, ALP1-like protein, and two uncharacterized proteins (Table 2). Besides genes directly tagged by DArT-seq marker alignments, some neighboring sequences that could be present in the same LD blocks were also chosen. Those included putative transcription factors from bZIP, MYB-HB-like, C2H2, MADS-MIKC and WD40-like families, Zf-FLZ domain-containing protein, replication termination factor 2 RTF2, small and long non-coding RNAs, mir-393 precursor and other putative regulatory sequences. Their potential involvement in flowering regulatory pathways and putative association with flowering induction is presented in Discussion. Finally, we conducted an analysis of microsynteny shared between *L. luteus* and *L. albus* genome assemblies in regions adjacent to the selected significantly associated DArT markers. The lack of microsynteny between *L. albus* and *L. luteus* was observed only for the genome region carrying M092420 marker sequence, where two neighboring genes located *L. albus* chromosome 25, *Lalb_Chr25g0286361* and *Lalb_Chr25g0286351*, were mapped to loci localized on two different chromosomes in *L. luteus*, namely LLUT_LOCUS27614 on chromosome YL-19 and LLUT_LOCUS11214 on chromosome YL-07. Moreover, an *L. luteus* LLUT_LOCUS18135 gene on chromosome YL-12 carrying M089187 marker sequence aligned to two genes in *L. albus* genome, *Lalb_Chr06g0173461* and *Lalb_Chr21g0316521*. The latter one revealed higher sequence similarity to LLUT_LOCUS2261 on different chromosome (YL-15) to the original LLUT_LOCUS18135 (Table 2).

Discussion

DArT sequencing is an effective method for genotyping non-model plants

Diversity Array Technology, DArT, is a sequence-independent, microarray-based, a high-throughput genome analysis method that enables multiplexing and allows the simultaneous interrogation of several hundred to several thousand polymorphic loci across the genome. This type of polymorphism is based on sequence variations and methylation patterns at the recognition sites of restriction enzymes (REs) and insertions or deletions (INDELS).

Marker	<i>L. luteus</i> gene	<i>L. albus</i> gene	Annotation
M104942	N/D	<i>Lalb_Chr14g0365811*</i>	non-coding RNA
M104942	N/D	Cluster_73519	non-coding RNA
M104942	N/D	<i>Lalb_Chr14g0365831</i>	Putative transcription factor bZIP family/TGA2.3-like
M079035	LLUT_LOCUS32250*	<i>Lalb_Chr09g0329621*</i>	Putative Zf-FLZ domain-containing protein
M079035	N/D	Cluster_49371	long non-coding RNA
M102006	LLUT_LOCUS31495*	<i>Lalb_Chr13g0300951*</i>	Putative pre-mRNA polyadenylation factor Fip1
M102006	LLUT_LOCUS31494	<i>Lalb_Chr13g0300941</i>	Putative Ubiquitin domain, replication termination factor 2 RTF2, Zinc finger, RING/FYVE/PHD-type
M074938	LLUT_LOCUS17529*	<i>Lalb_Chr02g0146461*</i>	Putative pectinesterase/pectinesterase inhibitor 41
M055339	LLUT_LOCUS4667*	<i>Lalb_Chr03g0032501*</i>	Unknown/hypothetical protein
M055339	LLUT_LOCUS4666	<i>Lalb_Chr03g0032521</i>	Putative transcription factor bZIP/ABA INSENSITIVE 5-like protein 2
M055339	LLUT_LOCUS4656	<i>Lalb_Chr03g0032571</i>	Putative transcription factor MYB-HB-like family/GAMBYB-like
M079300	LLUT_LOCUS36319*	<i>Lalb_Chr01g0012161*</i>	Putative ribosomal protein L34Ae
M079300	LLUT_LOCUS36317	<i>Lalb_Chr01g0012131</i>	Putative transcription factor C2H2 family/zinc finger protein 1-like
M065429	LLUT_LOCUS34632*	<i>Lalb_Chr19g0132221*</i>	Putative cellulose synthase (UDP-forming) chromatin regulator PHD family
M065429	LLUT_LOCUS34631	<i>Lalb_Chr19g0132201</i>	Putative transcription factor C2H2 family/E3 ubiquitin ligase BIG BROTHER
M089187	LLUT_LOCUS18135*	<i>Lalb_Chr06g0173461*</i>	Transcription initiation factor IIB
M089187	LLUT_LOCUS22611	<i>Lalb_Chr21g0316521</i>	Putative transcription factor MADS-MIKC family/ Agamous-like MADS-box protein AGL18
N114410	LLUT_LOCUS23544*	<i>Lalb_Chr14g0368451*</i>	Putative armadillo-like helical, pumilio, RNA binding domain-containing protein
N114410	LLUT_LOCUS23547	<i>Lalb_Chr14g0368401</i>	Putative transcription factor MYB-HB-like family/MYB44-like
M076759	LLUT_LOCUS17529*	<i>Lalb_Chr02g0146461*</i>	Putative pectinesterase/pectinesterase inhibitor 41
M103926	LLUT_LOCUS4442*	<i>Lalb_Chr03g0034641*</i>	Putative cytochrome P450/MFP1 attachment factor
M103926	N/D	<i>Lalb_Chr03g0034631</i>	mir-393 microRNA precursor family
M075705	LLUT_LOCUS5864*	<i>Lalb_Chr11g0074201*</i>	Putative DNA helicase/ATP-dependent DNA helicase Q-like 3
M075705	LLUT_LOCUS5876	<i>Lalb_Chr11g0074111</i>	Putative Zf-FLZ domain-containing protein/FCS-Like Zinc finger 13
M092420	LLUT_LOCUS27614*	<i>Lalb_Chr25g0286361*</i>	Putative protein-serine/threonine kinase CMGC-CDK-CRK7-CDK9 family
M092420	LLUT_LOCUS11214	<i>Lalb_Chr25g0286351</i>	Putative protein kinase PEK-PEK family transcription factor WD40-like family/SPA1-RELATED 4-like
M103410	LLUT_LOCUS13188*	<i>Lalb_Chr05g0219221*</i>	Antagonist of like heterochromatin protein 1-like

Table 2. Genetic annotation of selected DArT-seq markers, highlighted by GWAS as significantly associated. Two lupin genomes (*L. luteus* and *L. albus*) were used for BLAST mapping of DArT-seq markers. Loci directly matching DArT-seq markers are indicated by asterisks, whereas corresponding neighboring loci (if applicable) are listed in row(s) below them.

REs-based methods provide greater accuracy, resulting in improved reproducibility compared to PCR-based methods, which rely on the lower precision from selective primer binding^{52,53}. It was demonstrated that the vast majority of sequenced DArT markers (about 97%) scored identical, irrespective of chosen DNA preparations, but the remaining 3% gave consistently different results for different DNA samples. It suggests that a marginal effect of DNA methylation on the outcome needs to be considered⁵⁴. Studies on white lupin and narrow-leaved lupin have demonstrated that, although rearrangements in the promoter regions of *FT* genes have a significant impact, none of the DArT markers have been found near their sequences^{23,41,55,56}. This is likely due to the conserved nature of promoter regions, which results in a lack of suitable cleavage sites for restriction enzymes and selective pressure against polymorphisms in these areas. As a result, the likelihood of detecting DArT markers in promoter regions is lower compared to other parts of the genome, such as introns or intergenic regions. Nonetheless, the DArT markers have been integrated into analyses of many plant species, including major crops like wheat, barley, common bean, and rice, as well as non-model plants such as lupin^{23,54,57,58,59,60}. They have been employed extensively for the construction of molecular maps, identifying trait-marker associations, assessment of genetic diversity, association mapping, and routine genotyping in various crops for varietal identification^{60,61,62,63,64,65}.

Perspectives for adaptation of *L. hispanicus* for spring and autumn sowing in Europe

Greenhouse observations performed in the study revealed high variability of flowering time and vernalization responsiveness in *L. hispanicus*, highlighted by wide range of days to flowering of non-vernalized plants (from 79.3 to more than 160 days) and significant differences between accessions in acceleration of flowering by vernalization treatment (from 15.2 to more than 93 days). The number of days from sowing to flowering of non-vernalized late *L. hispanicus* accessions observed in the present study converges with previously reported values of 143, 164–183, and 211 days^{12,14,66}. Early and low-responsive to vernalization accessions of *L. hispanicus* were not reported previously to our knowledge. Georeferenced records localize *L. hispanicus* specimens in Iberian Peninsula regions with climatic conditions enabling effective vernalization during winter (<https://www.gbif.org/species/2964255>), including regular occurrence of mild frost (a few °C below zero) in many locations^{14,67,68}. A large number of *L. hispanicus* locations experience more than 1200 chilling hours during typical winter⁶⁹. A value of 1200 chilling hours corresponds to about 50 days of effective vernalization, well above the three-week period required for Old World lupins to be physiologically vernalized^{22,23,24,35,55}. Due to the very frequent occurrence of late spring and early summer droughts in the Iberian Peninsula climate, *L. hispanicus*, like many other plant species, evolved a drought escape strategy by autumn seed germination, winter growth, and significant spring phenological response to winter vernalization towards flowering and yield production⁷⁰. Such a strong response to vernalization has been demonstrated in our study by a much shorter range and lower values of growing degree days (GDDs) from sowing to the start of flowering with applied vernalization treatment (from 1018 to 1882 GDDs) than without it (from 1412 to 3211 GDDs). Similar experiments performed for *L. albus* revealed the values from sowing to start of flowering in the range from 698 to 2397 GDDs without vernalization and from 518 to 1435 GDDs with vernalization²³. These values are lower than those measured for *L. hispanicus*, however, vernalization responsiveness in some late flowering *L. albus* accessions was also as sound as in *L. hispanicus*. Based on the reported number of days from sowing to flowering, we were able to calculate GDDs also for two other Old World lupin crop species, *L. angustifolius* and *L. luteus*^{22,55}. This calculation revealed that *L. angustifolius* flowered in Poznań 716 to 2250 GDDs after sowing without vernalization and 640 to 1164 GDDs after vernalization treatment, whereas for *L. luteus* these values were 873–1816 GDDs and 763–1709 GDDs, respectively (Table 3). Comparison of those values confirmed reported much later phenology of *L. hispanicus* than all domesticated Old World lupin species and revealed its significantly higher vernalization responsiveness^{12,14,66}. Nevertheless, the earliest *L. hispanicus* lines could be placed in the middle of the GDDs range reported for lupin cultivars, intermediating between winter- and spring ecotype.

To verify if *L. hispanicus* GDDs values correspond to the climatic variables conditioning lupin spring sowing in western Poland, we calculated GDDs for 2022 and 2023 growing seasons in Poznań using meteorological data reported by the Institute of Meteorology and Water Management - National Research Institute in Poland (<https://danepubliczne.imgw.pl/pl>). Assuming spring sowing on the 1st April, the threshold of 2000 GDDs was reached on 24th August in 2022 and 18th August in 2023, whereas with sowing on 1st May (to avoid common spring frosts), this threshold was achieved on the beginning of September in both years^{71,72}. Therefore, expected performance of non-vernalized early *L. hispanicus* accessions in this location could converge with the moderately late accessions of soybean, yielding in middle autumn, when weather may become less favorable for pod drying⁷³. It should be noted that there is a significant SW-NE decreasing GDDs gradient in Poland, following transition from oceanic to humid continental climate⁷⁴. Taking into consideration the whole European continent, a clear latitudinal pattern of GDDs has been observed, with the highest values, recalculated to the base temperature of 3 °C, exceeding 4000 GDDs in the Mediterranean and the Black Sea regions⁷⁵. Therefore, the earliest *L. hispanicus* accessions should develop seed yield across a wide range of European environments, even if sown in April/May and cultivated in temperatures above vernalization threshold for the whole growing period, if properly watered.

On the contrary, intermediate and, particularly, late *L. hispanicus* accessions will have to be subjected to at least partial vernalization, even in the warmest regions of the continent, to ensure timely flowering and dry pod harvesting. It can be achieved by autumn sowing, however, artificial adaptation of *L. hispanicus* to harsh winters outside mild Mediterranean climate zone requires selection towards frost resistance. It should be some frost resistance that had already evolved in natural *L. hispanicus* populations, because plant annual communities growing on the semi-arid Mediterranean steppe in Spain exhibited very good performance until frost reaching –4.0 °C, whereas the lethal temperature causing 50% frost damage (LT50) was relatively low (about –9.5 °C)⁶⁸. Recent evaluation of *L. albus* frost tolerance using high-throughput phenotyping platform revealed that the most winter-adapted accessions have the 4-hours LT50 value about –12.0 °C, however, with significant visual

Species	Year	Vernalization	Environment	Mean GDDs	Min GDDs	Max GDDs
<i>L. hispanicus</i>	2022	—	CE (Poznań)	2410	1412	3177
	2022	+ +	CE (Poznań)	1262	1018	1882
	2023	—	CE (Poznań)	2362	1700	3211
	2023	+ +	CE (Poznań)	1376	1151	1744
<i>L. luteus</i>	2016	—	CE (Poznań)	1283	927	1816
	2016	+ +	CE (Poznań)	1183	918	1566
	2017	—	CE (Poznań)	1220	873	1745
	2017	+ +	CE (Poznań)	1078	777	1510
	2019	—	CE (Poznań)	1268	889	1742
	2019	+ +	CE (Poznań)	1116	763	1709
<i>L. albus</i>	2020	—	CE (Poznań)	1308	726	2397
	2021	—	CE (Poznań)	1229	698	2283
	2004	+ +	F (Lodi)	870	651	1048
	2005	+	F (S. Sauvant)	698	518	1100
	2004	+	F (Sanluri)	996	816	1435
<i>L. angustifolius</i>	2014	—	CE (Poznań)	1340	716	2250
	2014	+ +	CE (Poznań)	813	640	1164

Table 3. Comparison of the total number of growing degree days (GDDs) from sowing to flowering between *L. hispanicus* and three old world domesticated lupin species (*L. luteus*, *L. albus* and *L. angustifolius*) in greenhouse (CE) and field (F) conditions^{22,55}.

biomass injury starting from -9°C ⁷⁶. Preceding study encompassing all three Old World lupin crop species reported moderate frost tolerance of winter *L. albus* and *L. angustifolius* cultivars (up to -10°C , so well-aligned with both studies) and low frost tolerance of *L. luteus*, albeit without information about *L. luteus* ecotypes used in the study (spring or winter)^{22,34,77}. As vernalization responsiveness is highly correlated with frost tolerance in *L. albus*, it can be expected that late flowering *L. hispanicus* accessions that are highly responsive to vernalization will have higher frost tolerance than the early ones. Reaching the frost visual damage threshold value similar to that reported for winter ecotypes of *L. albus* probably will enable *L. hispanicus* yielding as a winter crop in the large part of the European continent⁷⁶.

The present study revealed significant association of *L. hispanicus* phenology with population structure, manifested by high correlations between assignment to particular clusters and the number of days from sowing to floral bud emergence, start of flowering and pod maturity. Similar observation was made for *L. albus* where early flowering genotypes from East African resources (Ethiopia, Kenya and Sudan), Egypt, Maghreb (Morocco) and West Asia (Israel and Syria) grouped together, whereas late flowering Azorean, Madeiran and Canarian genotypes also formed a separate cluster²³. Moreover, such diversification was also found in two other Old World domesticated species, *L. angustifolius* and *L. luteus*, where Palestinian accessions revealed distinct phenology^{24,78}.

Candidate genes controlling *L. hispanicus* phenology

Alignment of markers that revealed significant associations in BLINK or FarmCPU with at least four variables to the genomes of closely related species, *L. luteus* and *L. albus*, enabled us to search for candidate genes by analysis of microsyntenic blocks shared between all three species. The M102006 marker, especially sound in non-vernalized plants, aligned to the LLUT_LOCUS31495 gene on YL-22 and putative pre-mRNA polyadenylation factor Fip1 (Lalb_Chr13g0300951) gene on WL-13. The FIP1, Factor Interacting with Poly(A) Polymerase 1, a component of the pre-mRNA 3' end processing machinery and in *Arabidopsis* plays a significant role in the processes of seed dormancy and germination. FIP1 is predominantly expressed in seeds, and the knockout of FIP1 results in decreased seed dormancy and insensitivity to exogenous abscisic acid (ABA) during seed germination and the early stages of seedling establishment⁷⁹. Furthermore, the region of the WL-13 chromosome located ~ 100 kb from the current marker, was highlighted in a previous study to carry the Chr13_13913452_D marker significantly associated with flowering induction in *L. albus*²³. Moreover, the major QTL for flowering time identified in this species was also located here^{60,80,81,82}. One PAV marker from this region, Chr13_12561729_D, was successfully transformed to a PCR marker and validated in a large *L. albus* germplasm pool to be significantly associated with flowering time in the absence of vernalization⁶⁴.

An M079035 marker, that showed similar effects as M102006 marker (Fig. 4), was mapped to the LLUT_LOC32250 gene on YL-23 in *L. luteus* and to a gene encoding a putative Zf-FLZ domain-containing protein (Lalb_Chr09g0329621) on WL-09 in *L. albus*. BLAST analysis of both genes revealed considerable similarity to the transcription factor FCS-Like, which in *A. thaliana* exhibits high expression levels in reproductive organs, such as flowers. This observation may indicate its potential role in the regulation of organ development or transition between vegetative and reproductive phases⁸³.

The next marker, M074938 was the only marker significantly associated with non-vernalized plants in both 2022 and 2023. Moreover, the BLUE-based GWAS which combined 2022 and 2023 phenology revealed only one new marker, M076759 which demonstrated a remarkable similarity to M074938, both in terms of position on

L. luteus chromosome and allele segregation profile. Both these markers were mapped to loci on two, different chromosomes in *L. luteus* (YL-12 and YL-01) and *L. albus* (WL-06 and WL-02). All highlighted sequences demonstrated the highest similarity to the pectinesterase inhibitor 41 (PMEI41), which may play a role in post-transcriptional regulation. In transgenic *Arabidopsis* plants, under both long-day and short-day photoperiods, PME12 and PME14 facilitated early flowering phenotypes, accompanied by elevated expression levels of multiple flowering-related genes⁸⁴. Notably, in WL-02, this marker was located approximately 600 kb from two significant markers, Chr02_2625514_D and Chr02_2625564_D, which delimit the second major flowering time QTL in *L. albus*^{60,80,82}.

Another marker associated with non-vernalized plants flowering, M055339, has been identified in the YL-03 and WL-03 regions, with both instances classified as uncharacterized transcripts or hypothetical proteins. Notably, the adjacent genes in *L. luteus*, located 4 kb downstream (LLUT_LOCUS4666), and in *L. albus*, positioned 6 kb upstream (Lalb_Ch03g0032521), were annotated as 'Putative transcription factor bZIP family'. These genes exhibit the highest similarity to ABA INSENSITIVE 5-like protein 2, which plays a significant role in regulating plant growth and development⁸⁵.

An M103926 marker significant for non-vernalized plants was mapped with relatively low sequence identity of ~84% in YL-12 and WL-03. Both loci demonstrate similarity to Cytokinin hydroxylase-like/Putative cytochrome P450 gene. Notably, this hit is situated within a microRNA-rich region of *L. albus*, with the nearest microRNA cluster identified as the mir-393 microRNA precursor. In *Arabidopsis*, miR393a and miR393b target F-box auxin receptors TIR1 (TRANSPORT INHIBITOR RESPONSE 1), AFB1 (AUXIN SIGNALING F BOX PROTEIN 1), AFB2, and AFB3⁸⁶. The overexpression of miR393 led to a decrease in TIR1 mRNA levels; conversely, the miR393-resistant variant of TIR1 (mTIR1) caused an accumulation of mTIR1 transcripts, which reportedly induced delayed flowering in *Arabidopsis*, potentially by modifying the expression of various auxin-responsive genes⁸⁷.

The most significant BLAST results for marker M092420, which has been mapped to YL-19 and WL-25, were identified as members of the putative protein-serine/threonine kinase family. Notably, the closest gene, located approximately 1 kb downstream in *L. albus*, has been annotated as a putative protein kinase from the PEK-PEK family and resembles a WD40-like transcription factor. This gene exhibits the highest similarity to SPA-1-RELATED 4-like proteins in *Arabidopsis*, which are known to regulate photoperiodic flowering by interacting with CO to modulate its stability⁸⁸. Furthermore, this significant hit is situated within a previously characterized region of WL-25 associated with flowering time²³.

The last marker for flowering of non-vernalized plants that matched flowering-related genes by microsynteny, M103410, mapped to YL-09 and WL-05. A BLAST analysis of both locations revealed the highest similarity with ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1, ALP1. In *Arabidopsis*, ALP2, in conjunction with ALP1, demonstrates a reduced expression of floral identity genes such as PI, AG, and SEP3, indicating a role for ALPs in floral induction. Furthermore, protein interaction assays indicate that ALP2 binds directly with MSI1 and is essential for the interaction between ALP1 and PRC2⁸⁹.

The M104942 marker, associated with flowering time of vernalized plants, aligned to the non-coding region of *L. luteus* chromosome YL-16 and the *L. albus* chromosome WL-15 region carrying the long non-coding RNA cluster (Lalb_Ch14g0365811). The nearest neighboring gene, positioned ~2.5 kb upstream, is a putative transcription factor from the bZIP family (Lalb_Ch14g0365831), identified by BLAST analysis as the TGA2.3-like transcription factor. It has been already demonstrated that TGA transcription factors can influence flowering time, such as *Arabidopsis thaliana* AtTGA4 that directly interacts with the promoter region of CONSTANS (CO), a key regulator of the *Flowering locus T* (FT). Furthermore, the functional loss of AtTGA7 delayed *Arabidopsis* flowering by altered expression of the critical flowering inhibitor gene, *FLC*⁹⁰.

The following marker associated with flowering of vernalized plants, M079300, localized to LLUT_LOCUS36319 gene on YL-26 that did not match any *L. albus* gene with known involvement in regulation of flowering. However, the adjacent LLUT_LOCUS36318 gene, located 7 kb downstream, mapped in *L. albus* ~6 kb from a putative transcription factor from the C2H2 family, which is highly similar to Zinc Finger Protein 1-like (ZFP1-like). This protein, along with ZFP8, is known to be down-regulated by LEAFY and APETALA1, thereby alleviating their repression of class B and C floral homeotic genes⁹¹.

An M089187 marker, also associated with vernalized plant flowering, yielded one signal in *L. luteus* chromosome YL-12 and two hits in *L. albus*, localized on the WL-06 and WL-21 chromosomes. The hits corresponding to YL-12 and WL-06 exhibited a high degree of similarity with the Transcription Initiation Factor IIB. In contrast, the hit on WL-21 was located 10 kb from a Putative Transcription Factor of the MADS-MIKC family (Lalb_Ch21g0316521), highly similar to the AGAMOUS-like MADS-box protein AGL18. AGL18 functions as a transcription factor that plays a crucial role in the negative regulation of flowering via the photoperiodic pathway⁹².

Another marker that was significantly correlated for vernalized plants, M075705, mapped on YL-04 and WL-11 chromosomes, aligned to DNA helicases. Approximately 60 kilobases downstream from this marker on the WL-11 chromosome, there are two copies of putative Zf-FLZ domain-containing proteins (Lalb_Ch11g0074111 and Lalb_Ch11g0074121), identified as FLZ13. These proteins are known to interact with *FLC* and *ABI5*, thereby negatively regulating flowering time in *Arabidopsis*⁹³.

We were unable to identify good candidate genes for the markers M065429 and N114410, localized in *L. luteus* genome on chromosomes YL-25 and YL-16. It is likely that M065429 marker is situated in region that is not conserved between *L. hispanicus* and *L. luteus* genomes, because we found rapid LD decay around this locus (Fig. 7). It could explain the lack of hits to known genes involved in flowering regulation as there is putatively no microsynteny here. However, this explanation cannot be applied for the second marker, N114410, located in the middle of the large region with high LD values. The influence of factors beyond gene-mediated regulation, such as miRNA and long non-coding RNA (lncRNA) cannot be ruled out here.

Conclusions

The phenology of the earliest *L. hispanicus* accessions matches the middle of the range of growing degree days reported for existing Old World lupin cultivars, bridging the gap between winter and spring ecotypes. These early accessions are expected to perform similarly to moderately late soybean accessions, yielding in early to mid-autumn, exploiting the full length of the growing season, and progressively expanding in the era of changing climate. Given these factors, Spanish lupin may be considered a strong candidate for the development of new cultivars, providing a valuable addition to already cultivated species. It has the potential to produce significantly more pods on the main stem, outyielding current lupin crops. However, essential agronomic traits, such as disease resistance and alkaloid content, must be improved during the domestication process.

This study identified several molecular markers significantly associated with multiple observed traits. This finding was obtained through BLINK and FarmCPU GWAS and supported by LD analysis. These markers are located in microsyntenic blocks shared between yellow and white lupin species and carry genes associated with flowering induction, flower formation, vernalization response, and photoperiod control. Due to their association with a wide range of flowering time, the identified molecular markers can be utilized in future breeding efforts to select genetic material for developing new varieties adapted to local agroclimatic conditions.

Materials and methods

L. hispanicus germplasm diversity panel

The plant material encompassed 173 accessions obtained under the Standard Material Transfer Agreement (SMTA) from five gene banks: (1) El Centro de Investigaciones Científicas y Tecnológicas de Extremadura (CICYTEX), Guadajira, province of Badajoz, Spain; (2) Genebank Gatersleben of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gaterslaben, Germany; (3) United States Department of Agriculture, Agricultural Research Service, Western Regional Plant Introduction Station, Washington State University, Pullman, Washington, United States; (4) Australian Grains Genebank, The State of Victoria Department of Environment, Energy and Climate Action, Victoria, Australia; and (5) Poznan Plant Breeders Ltd., Wiatrowo Plant Breeding Station, Wiatrowo, Poland. The investigated germplasm diversity panel consisted of 107 wild representatives, 32 landraces, 3 cultivars, and 31 accessions with unknown status in the passport data. The list of accessions is provided in Supplementary Table S7.

Seeds vernalization

The vernalization was performed on all investigated accessions using the procedure developed for other lupin species⁸¹. In each phenotyping experiment, half of the seeds used were vernalized on Petri dishes with moist paper at 5 °C for 21 days in 2022 and 14 days in 2023. Vernalization was conducted in the darkness to prevent green mass and root development. Non-vernalized control plants were sown four days before the end of the vernalization procedure and grown at a temperature of ~22 °C (~10 °C above the vernalization threshold) to maintain a similar thermal time³¹.

Phenotyping of phonological phases

Greenhouse observations were performed during the 2022 and 2023 growing seasons. Phenotyping was performed at the Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland (52°26' N 16°54' E). Sowing dates for the non-vernalized plants were 13th March in 2022 and 10th March in 2023, while the vernalized plants were sown four days later (17th March in 2022 and 14th March in 2023). Automatic heating kept the minimum air temperature above 18 °C, whereas cooling was maintained by a temperature-dependent window-opening system (activated at 22 °C). The number of days from sowing to (a) the appearance of the first floral bud, (b) the appearance of the first fully colored petal (flowering), and (c) the date of harvesting mature pods from the main stem were recorded. Observations were made in three biological replicates every two days, from sowing to the harvesting of the last mature pods. The experimental design is provided in Supplementary Table S1.

The cumulative growing degree days (GDDs) were calculated using the formula

$$GDDs = \sum_{t=1}^n \max\left(\frac{T_{max} + T_{min}}{2} - T_b; 0\right)$$

Where t and n are days from sowing and the total number of days from sowing to the observed phenotypic trait (floral bud emergence, start of flowering, and pod maturity), T_{max} and T_{min} are daily maximum and minimum temperatures, whereas T_b represents the base temperature parameterized in this study for lupin as 3 °C^{34,94}. GDD values for fractional days were calculated on a linear scale.

Calculation of heritability

A statistical framework was employed using R and its advanced libraries. To account for spatial variability inherent in greenhouse data, linear mixed models (LMMs) were utilized. These models were implemented with the SpATS package, incorporating row and column effects as random factors and a spatial trend component to capture environmental heterogeneity within the greenhouse. Row and columns represent a two-dimensional arrangement of pots in the greenhouse. The spatial trend was estimated using the SAP function from SpATS, ensuring the model accurately reflected spatial variability. Heritability was estimated using a generalized broad-sense heritability formula which provides a robust measure of the proportion of phenotypic variance attributable to genetic factors

$$H^2 = \frac{\sum_{i=s+1}^m k_i}{m-s},$$

where k_i represents the eigenvalues derived from the genetic effects model, m denotes the total number of eigenvalues, and s indicates the number of eigenvalues that are zero due to model constraints⁹⁵. This approach accounts for both additive and non-additive genetic variances, offering a comprehensive insight into the genetic contributions to flowering traits. Analyses were conducted on datasets stratified by year to address potential temporal variability, ensuring results were reflective of annual differences.

DNA isolation

DNA was isolated from three biological replicates per every genotype. Each time, two young upper leaves (about 100 mg of tissue) were placed in 2 ml collection tubes (Eppendorf, Hamburg, Germany) and immediately frozen in liquid nitrogen. Homogenization of frozen plant tissue was performed using two stainless steel beads (ø 5 mm, Qiagen) placed inside the collection tubes and TissueLyser II homogenizer (Qiagen, Hilden, Germany) for 45 s at 30 rpm. DNA isolation was performed in automated station Maxwell® RSC 48 Instrument (Promega, Mannheim, Germany) using Maxwell® RSC PureFood GMO and Authentication Kit (Promega) according to the protocol provided by the manufacturer. DNA concentration and quality were estimated using a NanoDrop 2000 (ThermoFisher Scientific, Warsaw, Poland). The DNA concentration was ranging from 116.2 to 1910 ng/μl (mean value 896 ± 369 ng/μl). The ratio of absorbance at 260 nm vs. 280 nm ranged from 1.87 to 2.20 (mean value 2.08 ± 0.06), whereas at 260 nm vs. 230 nm from 1.46 to 2.33 (mean value 2.06 ± 0.13).

DArT-seq genotyping

DNA isolates from three biological replicates of every investigated genotype were diluted to 100 ng/μl and combined in equal aliquots to constitute bulked samples for DArTseq genotyping. The library preparation and the sequencing were conducted by Diversity Arrays Technology Pty Ltd. (University of Canberra, Bruce, Australia) using the Lupin DArTseq 1.0 protocol and high-density sequencing (2.5 mln reads). Obtained dataset included information about polymorphisms of the presence/absence variants (PAVs) of dominant markers (SilicoDArTs) and standard single nucleotide polymorphism (SNP) markers. To facilitate interpretation of the results, PAV markers were renamed to start with “M” letter whereas SNP markers to start with “N” letter.

Sequence data Preparation

The raw DArT data for SNP and PAV markers was imported using dartR package⁹⁶. Both files were subjected to four filtering steps, removing the following markers which: (a) were fully monomorphic in the given dataset (~ 231 000 markers), (b) showed the reproducibility rate below 97% (~ 600 markers) and (c) revealed the call rate by locus below 80% (~ 2000 markers). The output file was then submitted for the additional filtering using Microsoft Excel and following thresholds: maximum 20% of heterozygotes, minimum 5% of minor allele frequency (MAF) and, again, maximum 20% of missing genotype calls. The reference alleles were randomly nominated based on the order of the sequence obtained. The sequences of DArT-seq markers with trimmed adaptor regions were aligned to the *L. luteus* genome (GenBank BioProject PRJEB74252, assembly GCA_964019355.1) using BLAST algorithm implemented in Geneious software^{97,98}. The following parameters were applied: program blast n, word size 15, gap cost 5 (open) and 2 (extend), scoring 2 (match) and -3 (mismatch), e-value threshold 1×10^{-5} , maximum number of hits 1. BLAST results were further filtered for the minimum pairwise identity of 90% and the minimum query coverage of 75%. *L. luteus* genome coordinates of retained blast hits were used for GWAS as well as for searching of candidate genes.

Imputation of genotypes and population structure analysis

All marker data were transformed to 0, 1, 2 code, where 0 stands for the homozygote, 2 for the alternative allele homozygote, and 1 for the heterozygote. For data transformation, custom Python scripts were used. For missing data, imputation was performed using Beagle software version 4.1⁹⁹ with its default settings. Duplicated loci with identical segregation were removed, leaving a single representative. That prepared data was filtered for minor allele frequency (MAF) with the threshold of 5.0%. Population structure analysis was performed using the snmf function from the LAE package¹⁰⁰. This analysis estimated ancestral populations (K) in the range of 3 to 15, with 3000 replications and 5000 iterations for each K-value. The best run was determined using the cross-entropy criterion. For further analysis, population structure was visualized for the selected K=6 using the pophelper package, facilitating clear representation of the Q matrix and the underlying population structure¹⁰¹.

Genome-wide association mapping and visualization

GWAS was performed using two models: the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model¹⁰² and the Fixed and random model Circulating Probability Unification model (FarmCPU)¹⁰³ implemented in the GAPIT R package¹⁰⁴. In the analysis, we accounted for population structure (Q) through a Qmatrix and for relationships among individuals through a kinship (K) matrix¹⁰⁵, both using the marker data. Based on the cross-entropy analysis, the Qmatrix of covariates with K=6 was selected as the representative number of clusters for population structure in GWAS for each trait. The significance threshold for marker-trait associations (MTA) was set to $p=0.05$ after applying the false discovery rate (FDR)¹⁰⁶ correction. To control the False Discovery Rate (FDR) in genome-wide association analysis, the Benjamini-Hochberg (B-H) procedure was applied as implemented in the GAPIT package¹⁰⁴, following the method described by Benjamini and Hochberg¹⁰⁷.

To obtain genotype effect estimates across both years and vernalization treatments, we fitted a linear mixed model using the following structure:

$$Y = \mu + \text{Genotype}^{(f)} + \text{Year}^{(f)} + \text{Vernalization}^{(f)} + (\text{Genotype} \times \text{Vernalization})^{(f)} \\ + (\text{Genotype} \times \text{Year})^{(r)} + (\text{Year} : \text{Pot} : \text{Row})^{(r)} + (\text{Year} : \text{Pot} : \text{Column})^{(r)} + \epsilon$$

Where:

- **Genotype**: fixed effect (to extract unbiased BLUEs).
- **Year**: fixed effect (only two years observed).
- **Vernalization**: fixed effect (binary treatment: vernalized / non-vernalized).
- **Genotype × Vernalization**: fixed interaction term to model genotype-specific response to treatment.
- **Genotype × Year**: random interaction term to account for variation in genotype performance across years.
- **Row and Column** nested within **Pot** and **Year**: random effects to account for spatial heterogeneity in greenhouse conditions.
- ϵ : residual error.

The model was implemented in ASReml-R version 4.2 (VSN International, Hemel Hempstead, UK)¹⁰⁸. Genotype was modeled as a fixed effect to extract BLUEs (Best Linear Unbiased Estimates), which were used as phenotypic input for genome-wide association analysis. This modeling strategy was selected to avoid double shrinkage, which can occur if the genotype is treated as random in both the phenotypic model and the GWAS. As discussed by Bernal-Vasquez et al.¹⁰⁹, the use of BLUEs ensures unbiased estimation of genotype performance and is appropriate for two-stage GWAS pipelines. The mixed model also accounted for genotype × year interaction, treatment response (vernalization), and spatial heterogeneity in the greenhouse environment through nested row and column random effects.

Visualization of population structure, and violin plots was made in R software using packages ggplot2¹¹⁰, GAPIT¹⁰⁴, and pophelper¹¹¹. LD graphs were prepared using LDheatmap¹¹², whereas Manhattan plots in the qqman package (<https://cran.r-project.org/web/packages/qqman/index.html>).

To find candidate genes, *L. luteus* genome regions carrying selected DArT-seq markers significantly associated with analyzed traits (LLUT_LOCUS sequences or ~ 1000 bp fragments if no annotation was found at particular marker locus) were aligned to the *L. albus* genome using BLAST online tool available on the genome website [<https://www.whitelupin.fr/index.html>]. Alignments with the lowest e-value were considered as significant and used for selection of candidate genes based on annotation provided for matching *L. albus* loci on the website genome browser^{50,51}. Then, we conducted an analysis of the composition and arrangement of genes adjacent to the selected DArT markers, comparing gene order in the genome assemblies of *L. luteus* and *L. albus*. For genes lacking annotation or annotated as uncharacterized, we performed a BLAST search using NCBI and considered the annotation corresponding to the best match.

Data availability

The dataset generated and analyzed during the current study is available in the Zenodo repository (<https://zenodo.org>), at the link doi.org/10.5281/zenodo.15195286.

Received: 5 February 2025; Accepted: 24 April 2025

Published online: 02 May 2025

References

1. Sujak, A., Kotlarz, A. & Strobel, W. Compositional and nutritional evaluation of several lupin seeds. *Food Chem.* **98**, 711–719. <https://doi.org/10.1016/j.foodchem.2005.06.036> (2006).
2. Boschini, G., D'Agostina, A., Annicchiarico, P. & Arnoldi, A. The fatty acid composition of the oil from *Lupinus albus* Cv. Luxe as affected by environmental and agricultural factors. *Eur. Food Res. Technol.* **225**, 769–776. <https://doi.org/10.1007/s00217-006-0480-0> (2007).
3. Cooper, M. & Messina, C. D. Breeding crops for drought-affected environments and improved climate resilience. *Plant. Cell.* **35**, 162–186. <https://doi.org/10.1093/plcell/koac321> (2023).
4. Pecetti, L. et al. White lupin drought tolerance: genetic variation, trait genetic architecture, and Genome-Enabled prediction. *Int. J. Mol. Sci.* **24** <https://doi.org/10.3390/ijms24032351> (2023).
5. Matesanz, S. et al. Plasticity to drought and ecotypic differentiation in populations of a crop wild relative. *AoB PLANTS*. **12** <https://doi.org/10.1093/aobpla/plaa006> (2020).
6. Juzoń, K. et al. Polyamines in yellow lupin (*Lupinus luteus* L.) tolerance to soil drought. *Acta Physiol. Plant.* **39**, 202. <https://doi.org/10.1007/s11738-017-2500-z> (2017).
7. Burchardt, S. et al. Exploring the response of yellow lupine (*Lupinus luteus* L.) root to drought mediated by pathways related to phytohormones, lipid, and redox homeostasis. *BMC Plant Biol.* **24**, 1049. <https://doi.org/10.1186/s12870-024-05748-4> (2024).
8. Buirchel, B. J. & Cowling, W. A. in *Lupinus as crop plants: biology, production, and utilization* (eds J. S. Gladstones, C. A. Atkins, & J. Hamblin) Ch. 2, 41–66 CAB International, (1998).
9. Lesk, C. et al. Stronger temperature-moisture couplings exacerbate the impact of climate warming on global crop yields. *Nat. Food.* **2**, 683–691. <https://doi.org/10.1038/s43016-021-00341-6> (2021).
10. Lucas, M. M. et al. The future of lupin as a protein crop in Europe. *Front. Plant Sci.* **6** <https://doi.org/10.3389/fpls.2015.00705> (2015).
11. Lopez-Bellido, L. & Fuente, M. Lupin crop as an alternative source of protein. *Adv. Agron.* **40**, 239–295. [https://doi.org/10.1016/S0065-2113\(08\)60284-9](https://doi.org/10.1016/S0065-2113(08)60284-9) (1986).
12. Arrieta, V. & Besga, G. Descripción de poblaciones silvestres de *Lupinus hispanicus* Boiss. Et reuter. *Invest. Agrar. Prod. Prot. Veg.* 311–332. <https://doi.org/10.1007/s10722-022-01531-2> (1993).
13. Lema, M. & Soengas, P. Delve into the potential of wild populations of *Lupinus hispanicus*. (2022). <https://doi.org/10.21203/rs.3.rs-1789281/v1>
14. Lema, M. & Soengas, P. Identification of promising accessions of wild *Lupinus hispanicus* Boiss. Et reut. From Northwestern Spain. *Genet. Resour. Crop Evol.* **70**, 677–690. <https://doi.org/10.1007/s10722-022-01531-2> (2023).

15. Múzquiz, M., Burbano, C., Rey, C. & Cassinello, M. A chemical study of lupinus hispanicus seed—nutritional components. *J. Sci. Food. Agric.* **47**, 197–204. <https://doi.org/10.1002/jsfa.2740470207> (1989).
16. Arrieta, V., Besga, G. & Cordero, S. Seed coat permeability and its inheritance in a forage lupin (*Lupinus hispanicus*). *Euphytica* **75**, 173–177. <https://doi.org/10.1007/BF00025601> (1994).
17. Múzquiz, M., Burbano, C., Gorospe, M. J. & Ródenas, I. A chemical study of lupinus hispanicus seed—toxic and antinutritional components. *J. Sci. Food. Agric.* **47**, 205–214. <https://doi.org/10.1002/jsfa.2740470208> (1989).
18. Hondelmann, W. The lupin — ancient and modern crop plant. *Theor. Appl. Genet.* **68**, 1–9. <https://doi.org/10.1007/BF00252301> (1984).
19. Świecicki, W., Świecicki, W. & Nijaki, T. *Lupinus X hispanicolutes* - An interspecific hybrid of old world lupins. *Acta Soc. Bot. Pol.* **68**, 217–220. <https://doi.org/10.5586/asbp.1999.029> (1999).
20. Tripathy, K. P., Mukherjee, S., Mishra, A. K., Mann, M. E. & Williams, A. P. Climate change will accelerate the high-end risk of compound drought and heatwave events. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2219825120. <https://doi.org/10.1073/pnas.2219825120> (2023).
21. Hermann, M., Wernli, H. & Röthlisberger, M. Drastic increase in the magnitude of very rare summer-mean vapor pressure deficit extremes. *Nat. Commun.* **15**, 7022. <https://doi.org/10.1038/s41467-024-51305-w> (2024).
22. Plewiński, P. et al. FLOWERING LOCUS T indel variants confer vernalization-independent and photoperiod-insensitive flowering of yellow lupin (*Lupinus luteus* L.). *Hortic. Res.* **9**. <https://doi.org/10.1093/hr/uhac180> (2022).
23. Rychel-Bielska, S. et al. A GWAS study highlights significant associations between a series of indels in a FLOWERING LOCUS T gene promoter and flowering time in white lupin (*Lupinus albus* L.). *BMC Plant Biol.* **24**, 722. <https://doi.org/10.1186/s12870-024-05438-1> (2024).
24. Rychel-Bielska, S., Plewiński, P., Kozak, B., Galek, R. & Książkiewicz, M. Photoperiod and vernalization control of Flowering-Related genes: A case study of the Narrow-Leafed lupin (*Lupinus angustifolius* L.). *Front. Plant Sci.* **11**, 572135. <https://doi.org/10.3389/fpls.2020.572135> (2020).
25. Taylor, C. M. et al. INDEL variation in the regulatory region of the major flowering time gene LanFTc1 is associated with vernalisation response and flowering time in narrow-leaved lupin (*Lupinus angustifolius* L.). *Plant, cell & environment*, (2018). <https://doi.org/10.1111/pce.13320>
26. Minoli, S., Jägermeyr, J., Asseng, S., Urfels, A. & Müller, C. Global crop yields can be lifted by timely adaptation of growing periods to climate change. *Nat. Commun.* **13**, 7079. <https://doi.org/10.1038/s41467-022-34411-5> (2022).
27. Rezaei, E. E., Siebert, S., Hüging, H. & Ewert, F. Climate change effect on wheat phenology depends on cultivar change. *Sci. Rep.* **8**, 4891. <https://doi.org/10.1038/s41598-018-23101-2> (2018).
28. Zhang, T. et al. Climate change May outpace current wheat breeding yield improvements in North America. *Nat. Commun.* **13**, 5591. <https://doi.org/10.1038/s41467-022-33265-1> (2022).
29. Wu, X., Liu, H., Li, X., Tian, Y. & Mahecha, M. D. Responses of winter wheat yields to Warming-Mediated vernalization variations across temperate Europe. *Front. Ecol. Evol.* **5**. <https://doi.org/10.3389/fevo.2017.00126> (2017).
30. Lu, X. et al. Winter warming post floral initiation delays flowering via bud dormancy activation and affects yield in a winter annual crop. *Proc. Natl. Acad. Sci.* **119**, e2204355119. <https://doi.org/10.1073/pnas.2204355119> (2022).
31. HUYGHE, C. Winter growth of Autumn-sown white lupin (*Lupinus albus* L.) main apex growth model. *Ann. Botany.* **67**, 429–434. <https://doi.org/10.1093/oxfordjournals.aob.a088178> (1991).
32. Shield, I. F., Scott, T., Stevenson, H. J., Leach, J. E. & Todd, A. D. The causes of over-winter plant losses of autumn-sown white lupins (*Lupinus albus*) in different regions of the UK over three seasons. *J. Agricultural Sci.* **135**, 173–183. <https://doi.org/10.1017/S002185969008047> (2000).
33. Annicchiarico, P. & Iannucci, A. Winter survival of pea, faba bean and white lupin cultivars in contrasting Italian locations and sowing times, and implications for selection. *J. Agricultural Sci.* **145**, 611–622. <https://doi.org/10.1017/S0021859607007289> (2007).
34. Huyghe, C. & Papineau, J. Winter development of autumn sown white Lupin: agronomic and breeding consequences. *Agronomie* **10**, 709–716. <https://doi.org/10.1051/agro:19900902> (1990).
35. Adhikari, K. N., Buirchell, B. J. & Sweetingham, M. W. Length of vernalization period affects flowering time in three lupin species. *Plant. Breed.* **131**, 631–636. <https://doi.org/10.1111/j.1439-0523.2012.01996.x> (2012).
36. Rahman, M. & Gladstones, J. Control of lupin flower initiation by vernalization, photoperiod and temperature under controlled environment. *Aust. J. Exp. Agric.* **12**, 638–645. <https://doi.org/10.1071/EA9720638> (1972).
37. Annicchiarico, P., Harzic, N. & Carroni, A. M. Adaptation, diversity, and exploitation of global white lupin (*Lupinus albus* L.) landrace genetic resources. *Field Crops Res.* **119**, 114–124. <https://doi.org/10.1016/j.fcr.2010.06.022> (2010).
38. Cao, D. Vernalization or devernialization? A question about VRT2. *Plant Physiol.* **196**, 2266–2268. <https://doi.org/10.1093/plphys/kiae440> (2024).
39. Alkemade, J. A. et al. Genome-wide association study reveals white lupin candidate gene involved in anthracnose resistance. *Theor. Appl. Genet.* **135**, 1011–1024. <https://doi.org/10.1007/s00122-021-04014-7> (2022).
40. Yang, H. et al. Draft genome sequence, and a Sequence-Defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PLoS ONE*, **8**, e64799. <https://doi.org/10.1371/journal.pone.0064799> (2013).
41. Hane, J. K. et al. A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. *Plant Biotechnol. J.* **15**, 318–330. <https://doi.org/10.1111/pbi.12615> (2017).
42. Kozak, B., Galek, R., Zalewski, D. & Sawicka-Sienkiewicz, E. Preliminary genetic map of a new Recombinant inbred line population for Narrow-leaved lupin (*Lupinus angustifolius* L.). *Agronomy* **9**, 653. <https://doi.org/10.3390/agronomy9100653> (2019).
43. Plewinski, P., Książkiewicz, M., Rychel-Bielska, S., Rudy, E. & Wolko, B. Candidate Domestication-Related genes revealed by expression quantitative trait loci mapping of Narrow-Leafed lupin (*Lupinus angustifolius* L.). *Int. J. Mol. Sci.* **20**. <https://doi.org/10.3390/ijms20225670> (2019).
44. Annicchiarico, P. et al. White Lupin Adaptation to Moderately Calcareous Soils: Phenotypic Variation and Genome-Enabled Prediction. *Plants* **12**, 1139, (2023). <https://doi.org/10.3390/plants12051139>
45. Iqbal, M. M. et al. The first genetic map for yellow lupin enables genetic dissection of adaptation traits in an orphan grain legume crop. *BMC Genet.* **20**, 68. <https://doi.org/10.1186/s12863-019-0767-3> (2019).
46. Clauw, P., Ellis, T. J., Liu, H. J. & Sasaki, E. Beyond the standard GWAS—A guide for plant biologists. *Plant Cell Physiol.* <https://doi.org/10.1093/pcp/pcae079> (2024).
47. Schwertfirm, G. et al. Genome-wide association study revealed significant SNPs for anthracnose resistance, seed alkaloids and protein content in white lupin. *Theor. Appl. Genet.* **137**, 155. <https://doi.org/10.1007/s00122-024-04665-2> (2024).
48. Gulisano, A., Dechesne, A., Paulo, M. J. & Trindade, L. M. Investigating the potential of Andean lupin as a lignocellulosic feedstock for Europe: first genome-wide association study on *Lupinus mutabilis* biomass quality. *GCB Bioenergy.* **15**, 38–57. <https://doi.org/10.1111/gcbb.13006> (2023).
49. Martinez-Hernandez, J. E. & Levicoy, D. Haroldo Salvo-Garrido Chromosome-level genome assembly of yellow lupin (*Lupinus luteus*) provides novel insights into genome evolution, crop adaptation and seed protein in the three most cultivated lupins, PREPRINT (Version 1) available at research square. (2024). <https://doi.org/10.21203/rs.3.rs-4171664/v1>
50. Hufnagel, B. et al. Pangenome of white lupin provides insights into the diversity of the species. *Plant Biotechnol. J.* **19**, 2532–2543. <https://doi.org/10.1111/pbi.13678> (2021).

51. Hufnagel, B. et al. High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nat. Commun.* **11**, 492. <https://doi.org/10.1038/s41467-019-14197-9> (2020).
52. Jaccoud, D., Peng, K., Feinstein, D. & Kilian, A. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* **29**, e25–e25. <https://doi.org/10.1093/nar/29.4.e25> (2001).
53. James, K. E. et al. Diversity arrays technology (DART) for pan-genomic evolutionary studies of non-model organisms. *PLoS One*. **3**, e1682. <https://doi.org/10.1371/journal.pone.0001682> (2008).
54. Wenzl, P. et al. Diversity arrays technology (DART) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9915–9920. <https://doi.org/10.1073/pnas.0401076101> (2004).
55. Nelson, M. N. et al. The loss of vernalization requirement in narrow-leaved lupin is associated with a deletion in the promoter and de-repressed expression of a *Flowering locus T* (FT) homologue. *New Phytol.* **213**, 220–232. <https://doi.org/10.1111/nph.14094> (2017).
56. Książkiewicz, M. et al. Expansion of the phosphatidylethanolamine binding protein family in legumes: a case study of *Lupinus angustifolius* L. *FLOWERING LOCUS T* homologs, *LanFTc1* and *LanFTc2*. *BMC Genom.* **17** <https://doi.org/10.1186/s12864-016-3150-z> (2016).
57. Sansaloni, C. et al. Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nat. Commun.* **11** <https://doi.org/10.1038/s41467-020-18404-w> (2020).
58. Nemli, S., Aşçıoğlu, T., Ates, D., Eşiyok, D. & Tanyolac, B. Diversity and genetic analysis through DARTseq in common bean (*Phaseolus vulgaris* L.) germplasm from Turkey. Vol. 41 (2017).
59. Mogga, M., Sibiya, J., Shimelis, H., Lamo, J. & Yao, N. Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (*Oryza sativa* L.) using dart markers. *PLoS One*. **13**, e0198012. <https://doi.org/10.1371/journal.pone.0198012> (2018).
60. Vipin, C. A. et al. Construction of integrated linkage map of a Recombinant inbred line population of white lupin (*Lupinus albus* L.). *Breed. Sci.* **63**, 292–300. <https://doi.org/10.1270/jsbbs.63.292> (2013).
61. Vu, T. T. H., Kilian, A., Carling, J. & Lawn, R. J. Consensus genetic map and QTLs for morphological and agronomical traits in Mungbean using dart markers. *Physiol. Mol. Biology Plants: Int. J. Funct. Plant. Biology.* **28**, 1277–1295. <https://doi.org/10.1007/s12298-022-01201-8> (2022).
62. Wenzl, P. et al. A high-density consensus map of barley linking dart markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genom.* **7** <https://doi.org/10.1186/1471-2164-7-206> (2006).
63. Grynia, J. et al. The use of high-throughput DARTseq-based silicodart and SNP markers to identify yellow rust resistance genes in common wheat. *Euphytica* **220** <https://doi.org/10.1007/s10681-024-03291-9> (2024).
64. Surma, A. et al. Development and validation of PCR marker array for molecular selection towards spring, vernalization-independent and winter, vernalization-responsive ecotypes of white lupin (*Lupinus albus* L.). *Sci. Rep.* **15**, 2659. <https://doi.org/10.1038/s41598-025-86482-1> (2025).
65. Mangal, V. et al. Triumphs of genomic-assisted breeding in crop improvement. *Heliyon* **10**, e35513. <https://doi.org/10.1016/j.heliyon.2024.e35513> (2024).
66. Cowling, W. in *Plant Genetic Resources of Legumes in the Mediterranean* (eds Nigel Maxted & Sarita Jane Bennett) 191–206 Springer Netherlands, (2001).
67. García-Martín, A., Paniagua, L. L., Moral, F. J., Rebollo, F. J. & Rozas, M. A. Spatiotemporal analysis of the Frost regime in the Iberian Peninsula in the context of climate change (1975–2018). *Sustainability* **13**, 8491. <https://doi.org/10.3390/su13158491> (2021).
68. Pescador, D. S., Sánchez, A. M., Luzuriaga, A. L., Sierra-Almeida, A. & Escudero, A. Winter is coming: plant freezing resistance as a key functional trait for the assembly of annual mediterranean communities. *Ann. Bot.* **121**, 335–344. <https://doi.org/10.1093/aob/mcx166> (2018).
69. Moral, F. J., García-Martín, A., Rebollo, F. J., Rozas, M. A. & Paniagua, L. L. GIS-Based Analysis and Mapping of the Winter Chilling Hours in Mainland Spain. Application to Some Sweet Cherry Cultivars. *Agronomy* **11**, 330, (2021). <https://doi.org/10.3390/agronomy11020330>
70. Trullenque-Blanco, V., Beguería, S., Vicente-Serrano, S. M., Peña-Angulo, D. & González-Hidalgo, C. Catalogue of drought events in Peninsular Spanish along 1916–2020 period. *Sci. Data.* **11** (703). <https://doi.org/10.1038/s41597-024-03484-w> (2024).
71. Tomczyk, A. M., Szyga-Pluta, K. & Bednorz, E. Occurrence and synoptic background of strong and very strong Frost in spring and autumn in central Europe. *Int. J. Biometeorol.* **64**, 59–70. <https://doi.org/10.1007/s00484-019-01793-z> (2020).
72. Nidzgorska-Lenciewicz, J., Małkosa, A., Koźmiński, C. & Michalska, B. Potential risk of Frost in the growing season in Poland. *Agriculture* **14**, 501. <https://doi.org/10.3390/agriculture14030501> (2024).
73. Rychel-Bielska, S. et al. Molecular selection of soybean towards adaptation to central European agroclimatic conditions. *J. Appl. Genet.* <https://doi.org/10.1007/s13353-024-00889-6> (2024).
74. Wypych, A., Sulikowska, A., Ustrnul, Z. & Czekierda, D. Variability of growing degree days in Poland in response to ongoing climate changes in Europe. *Int. J. Biometeorol.* **61**, 49–59. <https://doi.org/10.1007/s00484-016-1190-3> (2017).
75. Spinoni, J., Vogt, J. & Barbosa, P. European degree-day climatologies and trends for the period 1951–2011. *Int. J. Climatol.* **35**, 25–36. <https://doi.org/10.1002/joc.3959> (2015).
76. Frangueli, N., Cavalli, D., Notario, T., Pecetti, L. & Annicchiarico, P. Frost tolerance improvement in pea and white lupin by a high-throughput phenotyping platform. *Front. Plant Sci.* **15**, 1490577. <https://doi.org/10.3389/fpls.2024.1490577> (2024).
77. Joernsgaard, B. & Kuptsov, N. in *Mexico, where old and new world lupins meet, the 11th International Lupin Conference*. 110–111 (International Lupin Association, Proceedings).
78. Kazimierski, T. & Kazimierska, M. Morphological and cytological differences within the species lupin *Us luteus* L. *Acta Soc. Bot. Pol.* **44**, 265–275. <https://doi.org/10.5586/asbp.1975.023> (2015).
79. Li, Y. et al. The Arabidopsis pre-mRNA 3' end processing related protein FIP1 promotes seed dormancy via the DOG1 and ABA pathways. *Plant. Journal: Cell. Mol. Biology.* **115**, 494–509. <https://doi.org/10.1111/tpj.16239> (2023).
80. Phan, H. T. T., Ellwood, S. R., Adhikari, K., Nelson, M. N. & Oliver, R. P. The first genetic and comparative map of white lupin (*Lupinus albus* L.): identification of QTLs for anthracnose resistance and flowering time, and a locus for alkaloid content. *DNA Research: Int. J. Rapid Publication Rep. Genes Genomes.* **14**, 59–70. <https://doi.org/10.1093/dnares/dsm009> (2007).
81. Rychel-Bielska, S. et al. Quantitative control of early flowering in white lupin (*Lupinus albus* L.). *Int. J. Mol. Sci.* **22**, 3856. <https://doi.org/10.3390/ijms22083856> (2021).
82. Książkiewicz, M. et al. A high-density consensus linkage map of white lupin highlights synteny with narrow-leaved lupin and provides markers tagging key agronomic traits. *Sci. Rep.* **7**, 15335. <https://doi.org/10.1038/s41598-017-15625-w> (2017).
83. Chen, S. et al. Genome-wide identification and characterization of FCS-Like zinc finger (FLZ) family genes in maize (*Zea mays*) and functional analysis of ZmFLZ25 in plant abscisic acid response. *Int. J. Mol. Sci.* **22**, 3529. <https://doi.org/10.3390/ijms22073529> (2021).
84. Li, B. et al. Genome-Wide identification of the PME1 gene family in tea plant and functional analysis of CsPME12 and CsPME14 through ectopic overexpression. *Front. Plant Sci.* **12** <https://doi.org/10.3389/fpls.2021.807514> (2022).
85. Vittozzi, Y., Krüger, T., Majee, A., Née, G. & Wenkel, S. ABI5 binding proteins: key players in coordinating plant growth and development. *Trends Plant Sci.* **29**, 1006–1017. <https://doi.org/10.1016/j.tplants.2024.03.009> (2024).
86. Jiang, J., Zhu, H., Li, N., Batley, J. & Wang, Y. The miR393-Target module regulates plant development and responses to biotic and abiotic stresses. *Int. J. Mol. Sci.* **23** <https://doi.org/10.3390/ijms23169477> (2022).

87. Teotia, S. & Tang, G. To bloom or not to bloom: role of MicroRNAs in plant flowering. *Mol. Plant*. **8**, 359–377. <https://doi.org/10.1016/j.molp.2014.12.018> (2015).
88. Laubinger, S. et al. Arabidopsis SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability. *Development* **133**, 3213–3222. <https://doi.org/10.1242/dev.02481> (2006).
89. Velanis, C. N. et al. The domesticated transposase ALP2 mediates formation of a novel polycomb protein complex by direct interaction with MSI1, a core subunit of polycomb repressive complex 2 (PRC2). *PLoS Genet.* **16**, e1008681. <https://doi.org/10.1371/journal.pgen.1008681> (2020).
90. Lu, C. et al. A comprehensive review of TGA transcription factors in plant growth, stress responses, and beyond. *Int. J. Biol. Macromol.* **258**, 128880. <https://doi.org/10.1016/j.ijbiomac.2023.128880> (2024).
91. Hu, T. et al. and APETALA1 down-regulate ZINC FINGER PROTEIN 1 and 8 to release their repression on class B and C floral homeotic genes. *Proceedings of the National Academy of Sciences* **120**, e2221181120, (2023). <https://doi.org/10.1073/pnas.2221181120>
92. Yan, K. et al. AGL18-1 delays flowering time through affecting expression of flowering-related genes in Brassica juncea. *Plant. Biotechnol. (Tokyo Japan)*. **35**, 357–363. <https://doi.org/10.5511/plantbiotechnology.18.0824a> (2018).
93. Li, X. et al. FLZ13 interacts with FLC and ABI5 to negatively regulate flowering time in Arabidopsis. *New Phytol.* **241**, 1334–1347. <https://doi.org/10.1111/nph.19445> (2024).
94. Huyghe, C. Winter growth of autumn-sown white lupin (*Lupinus albus* L.) main apex growth model. *Ann. Botany*. **67**, 429–434. <https://doi.org/10.1093/oxfordjournals.aob.a088178> (1991).
95. Oakey, H., Verbyla, A., Pitchford, W., Cullis, B. & Kuchel, H. Joint modeling of additive and non-additive genetic line effects in single field trials. *Theor. Appl. Genet.* **113**, 809–819. <https://doi.org/10.1007/s00122-006-0333-z> (2006).
96. Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C. & Georges, A. dartR v2: an accessible genetic analysis platform for conservation, ecology, and agriculture. *bioRxiv*, 2022.2003.2030.486475, (2022). <https://doi.org/10.1101/2022.03.30.486475>
97. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
98. Kears, M. et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> (2012).
99. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097. <https://doi.org/10.1086/521987> (2007).
100. Frichot, E. & François, O. L. E. A. An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929. <https://doi.org/10.1111/2041-210x.12382> (2015).
101. Francis, R. M. Pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **17**, 27–32. <https://doi.org/10.1111/1755-0998.12509> (2017).
102. Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8** <https://doi.org/10.1093/gigascience/giy154> (2019).
103. Liu, X., Huang, M., Fan, B., Buckler, E. S. & Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient Genome-Wide association studies. *PLoS Genet.* **12**, e1005767. <https://doi.org/10.1371/journal.pgen.1005767> (2016).
104. Lipka, A. E. et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444> (2012).
105. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423. 10.3168/jds.2007-0980 (2008).
106. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Ser. B (Methodological)*. **57**, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x> (1995).
107. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.: Ser. B (Methodol.)*. **57**, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x> (1995).
108. Butler, D., Cullis, B., Gilmour, A., Gogel, B. & Thompson, R. a. ASReml-R reference manual version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK, (2017).
109. Bernal-Vasquez, A. M. et al. The importance of phenotypic data analysis for genomic prediction - a case study comparing different Spatial models in Rye. *BMC Genom.* **15** <https://doi.org/10.1186/1471-2164-15-646> (2014).
110. Wickham, H. *ggplot2 Elegant Graphics for Data Analysis* (Springer, 2009).
111. Francis, R. M. Pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **17**, 27–32. <https://doi.org/10.1111/1755-0998.12509> (2017).
112. Shin, J. H., Blay, S., McNeney, B., Graham, J. & LDheatmap An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw. Code Snippets*. **16**, 1–9. <https://doi.org/10.18637/jss.v016.c03> (2006).

Acknowledgements

The authors would like to acknowledge: El Centro de Investigaciones Científicas y Tecnológicas de Extremadura (CICYTEX), Guadajira, province of Badajoz, Spain; Genebank Gatersleben of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany; United States Department of Agriculture, Agricultural Research Service, Western Regional Plant Introduction Station, Washington State University, Pullman, Washington, United States; Australian Grains Genebank, The State of Victoria Department of Environment, Energy and Climate Action, Victoria, Australia; and Poznan Plant Breeders Ltd., Wiatrowo Plant Breeding Station, Wiatrowo, Poland for providing the landrace accessions of Spain lupin. All seeds used were obtained under the terms of the SMTA (Standard Material Transfer Agreement) from the institutions mentioned above.

Author contributions

Conceptualization, W.B., S.R.-B.; methodology, W.B., S.R.-B.; software, W.B., and B.K.; validation, W.B., A.S., J.B., and S.R.-B.; formal analysis, W.B., B.K., and M.K.; investigation, W.B., A.S., J.B., and S.R.-B.; data curation, W.B., A.S., and S.R.-B.; writing—original draft preparation, W.B., M.K., and S.R.-B.; writing—review and editing, W.B., A.S., J.B., B.K., M.K. and S.R.-B.; visualization, W.B., B.K., and M.K.; supervision, W.B., M.K., S. R.-B.; project administration, W.B.; funding acquisition, W.B. All authors have read and agreed to the published version of the manuscript.

Funding sources

This study was funded by the National Science Centre, Poland (grant no. 2021/40/C/NZ9/00075) entitled “Molecular basis of the variability of flowering induction and response to vernalization in wild Old World lupins”. An open access charge was co-financed by Wrocław University of Environmental and Life Sciences as well as by National Science Centre, Poland. The funders had no role in the design of the study, in the collection,

analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Declarations

Competing interests

The authors declare no competing interests.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-00096-1>.

Correspondence and requests for materials should be addressed to S.R.-B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025, corrected publication 2025