



Harnessing historical genebank data to accelerate pea breeding

Lanique Niels¹ · Jochen Christoph Reif¹ · Lars-Gernot Otto¹ · Vilson Mirdita¹ · Markus Oppermann¹ · Ulrike Lohwasser¹ · Matthias Kotter¹ · Stephan Weise¹ · Samira El Hanafi¹

Received: 27 June 2025 / Accepted: 19 August 2025 / Published online: 9 September 2025
© The Author(s) 2025

Abstract

The German Federal Ex Situ Genebank for Agricultural and Horticultural Crops (IPK) harbours over 3000 pea plant genetic resources (PGRs), backed up by corresponding information across 16 key agronomic and economical traits. The unbalanced structure and inconsistent format of this historical data has precluded effective leverage of genebank accessions, despite the opportunities contained in its genetic diversity. Therefore, a three-step statistical approach founded in linear mixed models was implemented to enable a rigorous and targeted data curation. Spring accessions revealed considerable breeding potential, with protein content exceeding market standards by almost one-fifth and with hundred grain weight that could match the upper limits reported for European elite varieties. This variation is embedded within structured populations, comprising five convarieties including sugar snaps and field pea, adding value for breeding across diverse morphotypes and market segments. Winter accessions demonstrated cold resilience, with post-winter survival rate up to 79.27% under minimum temperatures as low as -17.1 °C. This variation is of particular relevance given the limited availability of winter-hardy cultivars able to evade summer drought and heat stresses. Transformation of the IPK Genebank into a bio-digital resource redirects formerly static material into central leverage for plant breeding in view of contemporary challenges. As such, this investigation activated the IPK pea population for use in among others breeding for a wide variety of ideotypes, research into adaptation, and future combination with omics studies.

Introduction

Despite being the backbone of global crop diversity conservation, most plant genetic resources (PGRs) remain poorly characterized and largely untapped in modern breeding. This disconnect persists even though searches of collections for traits of interest have identified suitable donor material in various cases, resulting in introgressions that addressed pressing issues (Guerra et al. 2022; Roorkiwal et al. 2020; Schulthess et al. 2022). Whereas genebank genetic diversity is plentiful and continues to offer the raw material needed to adapt crops to climate stress, market demands, and sustainability goals, the challenge lies in access, visibility and usability (Dempewolf et al. 2017; He et al. 2022; McCouch et al. 2013). Therefore, pre-breeders and breeders have often

turned to more narrowly defined breeding collections, comprising selected landraces and a few promising accessions from public collections (Engels et al. 2024). This reliance has intensified over time, particularly following the implementation of the Nagoya protocol, which, despite its role in promoting fairness, has introduced complexities in accessing and exchanging genetic resources. As a result, breeding programmes increasingly prioritize internally managed collections, while vast pools of diversity conserved in public genebanks, especially in underutilized crops, remain underexplored (Engels et al. 2024).

Pea (*Pisum sativum*), a high yielding temperate climate legume with valuable grain traits (Annicchiarico 2008; Carrouée et al. 2003; Crosta et al. 2023; Rogers et al. 2024; Thavarajah et al. 2023; Windsor et al. 2024), exemplifies this paradox. Despite its advantageous traits, it remains a minor crop due to marginal profitability compared to cereals (Magrini et al. 2016; Rubiales et al. 2021a) and soybean (FAO 2024; Rogers et al. 2024). Particularly, winter pea holds potential to increase relevance of pea cultivation due to its early sowing, season extension, and nitrogen savings, but is constrained by its sensitivity to frost and insufficient

Communicated by Matthew N Nelson.

✉ Samira El Hanafi
hanafi@ipk-gatersleben.de

¹ Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

winter hardiness. In a recent study, seedlings of elite genotypes demonstrated an expected 50% survival between -11.6 and -14.5 °C (Franguelli et al. 2024), exceeding commonly reported tolerated ranges between -8 and -13 °C (Liu et al. 2017; Urbatzka et al. 2012; Vogt-Kaute and Quendt 2022). Continuation and expansion of winter-hardiness is essential to enable reliable stand establishment and yield stability in cold-prone environment. Moreover, for spring and winter pea alike, development of more profitable varieties with the focus on yield, quality traits and resilience is essential, but is hampered by limited genetic diversity in the elite gene pools, resulting from decades of selection (Annicchiarico et al. 2017). The untapped genetic wealth of PGRs readily available in genebanks offers a potential solution (Qiu et al. 2013), provided we can overcome as a first step the bottlenecks in data usability. While many accessions have been phenotyped during regeneration periods, their associated data are often unbalanced, inconsistent, and unsuitable for direct use in breeding decisions (Anglin et al. 2018).

Therefore, curation of genebank data to meet the FAIR principles (Findable, Accessible, Interoperable, and Reusable; Wilkinson et al. 2016) is critical to transforming this underutilized wealth into an engine of crop innovation (Anglin et al. 2018). Historical data, once curated, can mimic long-term, multi-environment trials and explore trait plasticity, environmental adaptation, and rare variation. This potential has been demonstrated through initiatives at the IPK Genebank (González et al. 2018; Philipp et al. 2018), where curated records have enabled genomic studies to unlock new sources of variation and guide introgression strategies (Berkner et al. 2024b; El Hanafi et al. 2023; Jiang et al. 2021; Milner et al. 2019). Considering the broad variety of environmental and consumer demands, continuation and expansion of this trend is essential. Therefore, the search for new sources of diversity as input for research and breeding is as relevant as ever.

In line with this overarching need, the IPK Genebank's pea collection presents a valuable case study. As the sixth-largest and most diverse pea repository globally (Bagheri et al. 2023), it comprises over 5300 accessions including 109 winter pea accessions. Of these, respectively 3041 spring and 45 winter accessions are associated to detailed phenotype data, gathered from 76 and 7 countries. To overcome the limits imposed by its unbalanced historical data, advanced statistical methods are required to correct biases and extract meaningful patterns. Unlocking this dataset's full potential will support the development of climate-resilient and high-performing varieties, strengthening pea's role in sustainable agriculture. Therefore, the following objectives were to: (1) unlock the historical pea data of the IPK Genebank by utilizing a three-step data curation strategy; (2)

characterize phenotypic variation across spring and winter accessions; (3) identify frost-tolerant winter accessions for future breeding use; and (4) evaluate convarietal differentiation and trait interdependence to inform pre-breeding strategies.

Materials and methods

Plant material

This study focuses on 3086 pea accessions maintained at the German Federal Ex Situ Genebank at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany (Weise et al. 2025). Acquired through expeditions, seed exchanges, and breeding programmes, the collection spans a broad geographical range (Online Resource 1). Of these accessions, 2988 were classified as *Pisum sativum sativum* and include various convarieties according to Alefeld's classification: 1043 *sativum* (garden pea), 590 *speciosum* (field pea), 163 *axiphium* (snow pea), 900 *medullare* (wrinkled pea with high amylose content), and 30 *medullasaccharatum* (an intermediate form between *medullare* and *axiphium*; Alefeld 1866; Green 2008; Lehmann 1954; Osman and Ali 2021). In total 45 accessions were identified as winter type, characterized by a delayed growth when spring-sown and reliable winter survival after autumn sowing. From these, 38 belong to *P. sativum sativum speciosum* convariety. The passport data for these accessions are documented in the IPK Genebank Information System (GBIS; Oppermann et al. 2015).

Field trials

The phenotypic data were routinely collected at Gatersleben, Germany (latitude: $51^{\circ} 49' 22.5''$ N, longitude: $11^{\circ} 16' 40.6''$ E, 110.5 m.a.s.l.) during the seed regeneration process, which was carried out when seed stocks became insufficient or seed viability fell below a critical threshold. Regeneration trials followed an experimental design with no replication randomized across regeneration years, with a plot size of 2.88 m². The standard sowing rate was 100 seeds per plot, although limited seed availability occasionally required a reduced sowing rate. Additionally, in 2009, 2021, 2022, and 2023, a total of 9, 19, 29, and 22 spring accessions, respectively, were grown in Greenhouse facilities. The spring peas were sown between March and May, with data spanning from 1946 to 2023 (Online Resource 2). On average, each spring pea accession was evaluated in 4.36 years. Winter peas were sown in September or October, with most observations (84.72%) collected between 1955 and 1961. On average, each winter pea accession was tested in 4.8 years.

Phenotypic data were collected for 16 quantitative traits including phenological, agronomic, and yield-related traits (Table 1). Hundred grain weight (HGW) and protein content (Prot) of spring accessions were measured separately in a distinct trial setup in 1976. Accessions showing favourable performance for these two traits were regrown in 1978, and a subset was further evaluated in 1980 to assess performance across years (Müntz and Lehmann 1987).

Detailed environmental data were retrieved from a local weather station, including daily records of temperature, precipitation, and snow cover for the periods between November 1th to April 30th. Implausible minimum temperature values deviating by more than 20 °C from either adjacent days or by more than 15 °C from the corresponding minimum temperature at 2 m above ground were removed. Missing values were then imputed using the `na_kalman` function from the R package `imputeTS` (version 3.3; Moritz and Bartz-Beielstein 2017) or using a model adapted from Trnka et al. (2010) in case of snow cover with known precipitation.

Data analyses

Phenotypic data of spring and winter accessions was analysed separately. Plausibility checks were performed initially to exclude clearly erroneous observations. Observations falling outside of the range of the mean three times the standard deviation were excluded. Traits such as days to emergence (EmDays), start and end of flowering (FTS and FTE), and picking ripeness (Ripe) that exhibited unusually early or late values were flagged and discarded if they either (i) indicated an abnormal growing season, or (ii) were inconsistent with the patterns observed for the other three traits.

The following linear mixed model (LMM) was fitted to analyse the historical phenotypic data of each trait using the ASReml package in R (version 4.2.0.267; Butler et al. 2017; R Core Team 2024):

$$y_{ijkl} = \mu + g_i + a_j + t_k + v_l + g_i : t_k + e_{ijkl} \quad (1)$$

where y_{ijkl} is the phenotypic records for the i th accession tested in the j th year, within the k th experiment of the l th convariety. μ denotes the intercept, g_i the effect of the i th

Table 1 List of the 16 quantitative traits phenotyped by the IPK Gatersleben for the 3041 spring and 45 winter pea accessions, respectively.

Trait	Unit	Definition
Days to emergence (EmDays)	Days	Days at which 50% of the plants at the plot emerged, after March 1st for spring (<i>S</i>) or after sowing (<i>W</i>)
Start of flowering time (FTS)		Days after March 1st (<i>S</i>) or January 1st (<i>W</i>) at which the first flower opened for 50% of the plants in the plot
End of flowering time (FTE)		Days after March 1st (<i>S</i>) or January 1st (<i>W</i>) at which the last flower opened for 50% of the plants in the plot
Date of picking ripeness (Ripe)		Days after March 1st (<i>S</i>) or January 1st (<i>W</i>) at which 50% of the plants in the plot the first pod was suitable for commercial harvest
Duration of flowering (FTdur)		Number of days between the start and the end of flowering (<i>S</i> and <i>W</i>)
Plant height at flowering time (PHFT)	cm	Straight distance between the tip of the shoot and the soil, measured when 50% of the plants flowered and averaged across the plot (<i>S</i> and <i>W</i>)
Plant height (minimum) (PHmin)		Straight distance between the tip of the shoot and the soil of the shortest plant measured when 50% of the plot flowered, recorded in plots with accessions with variable heights (<i>S</i> and <i>W</i>)
Plant height (maximum) (PHmax)		Straight distance between the tip of the shoot and the soil of the tallest plant measured when 50% of the plot flowered, recorded in plots with accessions with variable heights (<i>S</i> and <i>W</i>)
Height of the first pod (PHpod)	cm	Height at the point where the stalk carrying the first pod branches off from the stem, averaged across different plants (<i>S</i> and <i>W</i>)
Number of flowers per inflorescence (Flr)	–	Classes indicating the number of flowers per inflorescence (1 = one, 2 = one or two, 3 = two, 4 = two or three, 5 = three or more), averaged across the plot (<i>S</i>)
Seed size (Ss)	–	Classes indicating the size of the seeds harvested from the entire plot (1 = small, 2 = medium, 3 = large; <i>S</i>)
Number of seeds per pod (Spp)	–	Average number of seeds per pod in 8 pods of 8 different plants (<i>S</i>)
Hundred grain weight (HGW)	g	Weight of one hundred grains, obtained by averaging the weight of 3 times 100 pea grains (<i>S</i> and <i>W</i>)
Protein content (Prot)	%	Protein content measured with the Kjeldahl method in air-dried pea grains (<i>S</i>)
Survival rate (Surv)		Number of vital plants observed in a row base within a plot, measured before and after winter (post-winter survival rate, PWSurv). Scored on 1–9 scale, and converted into percentage (<i>W</i>)

(*S*) and (*W*) indicate traits measured during spring and winter cultivation

accession, a_j the effect of the j th year, t_k the effect of the k th experiment, and v_l the effect of the l th convariety. The term $g_i:t_k$ captures the interaction between the genotype and experiment, and e_{ijkl} represents the residual. Full and reduced models were applied depending on the trait and the model convergence. v_l was omitted if its inclusion did not lead to significant improvement of the model fit.

Filtered data after plausibility check underwent a two-step outlier correction procedure. The first step involved year-based outlier detection to identify years with highly deviating values. Model 1 was fitted assuming accessions as random and years as fixed effects. Year effects and year-specific error effects were used to calculate the coefficient of variation (CV). If the CV for a given year exceeded three standard deviations from the mean CV, the corresponding trait-year records were removed. The second step involved record-based outlier correction, where model 1 was applied to the pre-corrected dataset from step one, with a slightly different assumption considering accessions as fixed and years as random effects. Standardized residuals were assessed, and any data points that fell beyond the Tukey-defined threshold were excluded (Bernal-Vasquez et al. 2016; Philipp et al. 2018).

Subsequently, Eq. (1) was re-fitted again to compute variance components and broad-sense heritability, as well as to calculate the best linear unbiased predictions (BLUEs). For BLUEs computation, the same assumptions in Eq. (1) as specified for outlier correction were considered, although convariety effects were omitted to retain differences in performance among accessions. Broad-sense heritability was computed assuming all factors as random effects except the intercept, with year and experiment combined into a single factor. To account for the unbalanced data, ad-hoc heritability was calculated as follows (Holand et al. 2003):

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gb}^2}{q} + \frac{\sigma_e^2}{p}}, \quad (2)$$

where $q = \sum_{i=1}^n \frac{g_i : b_i}{i}$ is the sum of each year x genotype interaction effect divided by their index, and $p = \sum_{i=1}^n \frac{e_{ijl}}{i}$ is the sum of each residual effect divided by their index. In the case of a reduced model, $\frac{\sigma_{gb}^2}{q}$ was omitted.

Convary effects on BLUEs were calculated by fitting:

$$y_{ijl} = \mu + v_l + e_{ijl} \quad (3)$$

Based on this model, the R package emmeans (version 1.10.4; Lenth 2024) was used to determine the significance level of pairwise differences with Holm's correction for multiple testing.

To assess the relationships between traits, Pearson's correlation coefficients for the resulting BLUEs of each trait pair within both winter and spring collections were further assessed using the R package WGCNA (1.72–5, Langfelder and Horvath 2008, 2012). Significance levels were adjusted with Holm's correction for multiple testing.

Results

Data curation improved the quality of the historical pea data

A three-step outlier correction strategy was implemented to enhance the quality of the IPK historical phenotypic data. As a result, an average of 1.3 and 0.2% of the accessions were excluded from the spring and winter populations across traits, respectively (Tables 2 and 3, Online Resources 3 and 4). Despite the relatively small proportion of the removed data, Heritability estimates increased substantially for some traits, particularly in the spring pea population, where estimates rose by up to 903% (Tables 2 and 3). This suggests that a small fraction of inconsistent data points had disproportionately inflated residual variance, obscuring meaningful differences in accession performance. In general, higher heritabilities were observed in the spring compared to the winter population, likely due to the larger number of spring accessions and trial years, which reduced environmental variability and improved performance estimates. Across both populations, phenology-related traits tended to show stronger genetic control, whereas certain traits in the winter population, such as Ripe and EmDays, exhibited very low heritability, highlighting the significant influence of environmental factors under winter conditions (Online Resource 5). Meanwhile, agronomic traits consistently displayed moderate-to-high heritability in both populations, indicating a consistent genetic contribution despite varying environmental effects.

Winter pea accessions exhibited substantial winterhardiness

From December to March, the minimum temperature measured at 5 cm above ground at Gatersleben location consistently remained below 0 °C across years, with February being on average the coldest month (Fig. 1a). Severe frost events, defined as temperatures below –14 °C with less than 5 cm snow cover, occurred in 7 out of the 10-year periods (Fig. 1b) in the timeframe between November 1st of the sowing year to April 30th in the following year. Due to the non-orthogonal design of the trials, winter pea accessions were exposed to varying levels of cold stress (Fig. 2, Online Resource 6). For example, accession PIS 402 experienced

Table 2 The estimated broad-sense heritability before and after the exclusion of outliers for each trait in the spring population

	N_{data}	$N_{outlier}$	N_{geno}	$N_{gen\sigma_{out}}$	Initial h^2	Final h^2	$\Delta h^2(\%)$
Days to emergence	11,653	280 (2.35%)	2941	49 (1.64%)	0.07	0.75	902.63
Start of flowering	11,436	277 (2.36%)	2936	54 (1.81%)	0.83	0.91	9.92
End of flowering	10,788	237 (2.15%)	2847	36 (1.25%)	0.68	0.79	16.64
Date of picking ripeness	9545	178 (1.83%)	2580	7 (0.27%)	0.68	0.81	18.55
Duration of flowering	10,704	186 (1.71%)	2863	10 (0.35%)	0.33	0.47	42.66
Plant height at flowering time	10,449	198 (1.86%)	2919	15 (0.51%)	0.87	0.90	3.3
Plant height minimal	687	46 (6.28%)	425	1 (0.23%)	0.83	0.92	10.42
Plant height maximal	797	63 (7.33%)	544	4 (0.73%)	0.84	0.93	11.54
Height of the first pod	8746	188 (2.1%)	2726	11 (0.4%)	0.78	0.83	6.41
Number of flowers per inflorescence	10,478	137 (1.29%)	2982	1 (0.03%)	0.69	0.74	7.52
Seed size	10,609	201 (1.86%)	2762	1 (0.04%)	0.72	0.77	7.1
Number of seeds per pod	8906	148 (1.63%)	2634	6 (0.23%)	0.52	0.60	14.3
Protein content	1719	167 (8.85%)	1283	70 (5.17%)	0.44	0.69	57.21
Hundred grain weight	1706	176 (9.35%)	1293	60 (4.43%)	0.85	0.92	8.95

Table 3 The estimated broad-sense heritability before and after the exclusion of outliers for each trait in the winter population

	N_{data}	$N_{outlier}$	N_{geno}	$N_{gen\sigma_{out}}$	Initial h^2	Final h^2	$\Delta h^2(\%)$
Days to emergence	313	3 (0.95%)	45	0 (0%)	0.00	0.00	0.98
Start of flowering	161	6 (3.59%)	42	0 (0%)	0.61	0.85	39.6
End of flowering	100	3 (2.91%)	38	0 (0%)	0.41	0.32	- 23.15
Date of picking ripeness	84	5 (5.62%)	38	0 (0%)	0.00	0.00	NA
Duration of flowering	99	3 (2.94%)	38	0 (0%)	0.00	0.21	NA
Plant height at flowering time	106	2 (1.85%)	35	0 (0%)	0.72	0.76	5.52
Plant height minimal	106	2 (1.85%)	35	0 (0%)	0.70	0.79	12.5
Plant height maximal	104	2 (1.89%)	35	0 (0%)	0.70	0.77	10.23
Height of the first pod	17	2 (10.53%)	11	0 (0%)	0.87	0.94	7.59
Hundred grain weight	116	2 (1.69%)	38	0 (0%)	0.75	0.85	13.16
Survival rate	159	5 (3.05%)	38	1 (2.56%)	0.00	0.11	NA
Post-winter survival rate	158	2 (1.25%)	39	0 (0%)	0.64	0.84	30.34

NA values indicate cases where no genotypic variance was detected, resulting in zero heritability

cold stress over 9 years, offering robust evidence of its consistent winterhardiness under recurrent extreme conditions. Notably, several accessions of the *axiphium* convariety (e.g. PIS 862) and *sativum* convariety (e.g. PIS 975 and PIS 32) survived at minimum temperature measured at 5 cm height below $-20\text{ }^{\circ}\text{C}$ with less than 5 cm snow cover (Fig. 3, Online Resource 7), highlighting their potential as valuable genetic resources for enhancing frost tolerance. Additionally, the *speciosum* convariety consistently demonstrated superior survival compared to others. These results surpass earlier thresholds reported from large-scale trials (-10 to $-18\text{ }^{\circ}\text{C}$; Liu et al. 2017; Zhang et al. 2016), from field trials with snow cover ($-20.3\text{ }^{\circ}\text{C}$; Homer et al. 2016) and from controlled settings (-8 to $-14.5\text{ }^{\circ}\text{C}$; Franguelli et al. 2024; Homer et al. 2016), underscoring the exceptional frost

tolerance documented in this study. The enhanced cold tolerance was independent of country of provenance, given that the site of collection or of evolutionary origin might differ. As such, survival rates were similarly high among accessions collected from diverse latitudes (e.g. Greece, Albania, and Germany).

Broad phenotypic variation to unlock trait-specific breeding improvement

To investigate the variation present within the IPK Genebank, BLUEs were calculated for each accession. The traits related to earliness, height, yield components, and post-winter survival rate (PWSurv) were prioritized for further analyses. Information of the remaining traits is

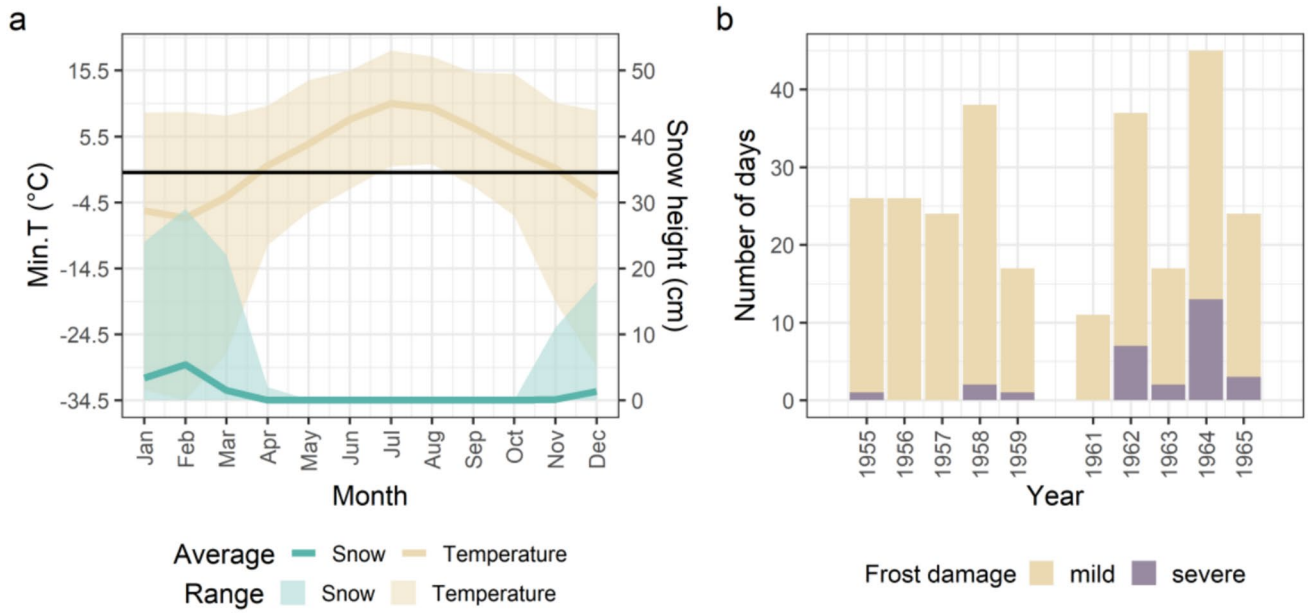


Fig. 1 Overview of the weather data during years with winter pea cultivation. **a** Monthly averages and ranges of the lowest daily temperatures (yellow) and snow cover (blue) measured 5 cm above ground from 1955 to 1965 (excluding 1960). **b** Number of days

per year with mild (minimum daily temperature at 5 cm above ground < - 6 °C) and severe (minimum daily temperature at 5 cm above ground < - 14 °C) frost stress under snow cover less than 5 cm

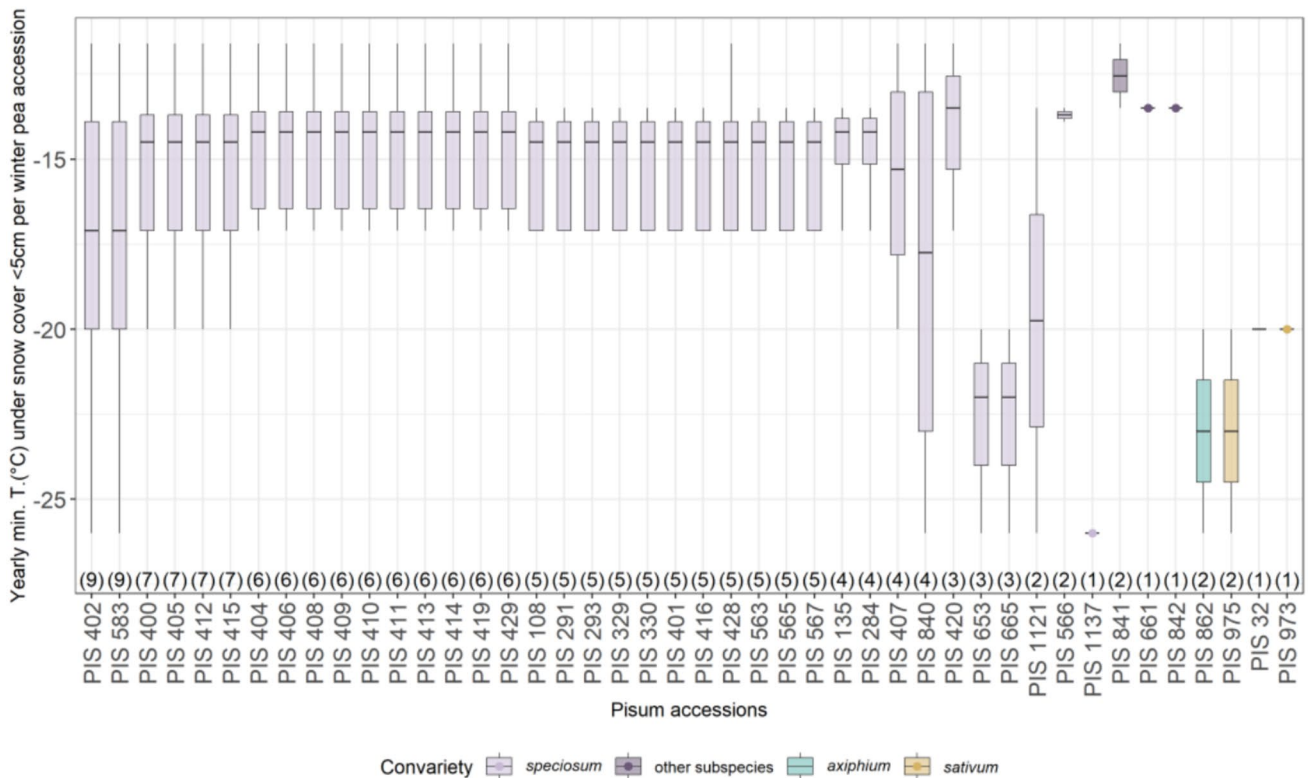


Fig. 2 Yearly minimum temperature at 5 cm above ground under snow cover < 5 cm for each winter pea accession. Numbers between brackets indicate the number of cultivation years per accession

presented in online resource 8. Distinct patterns emerged between spring and winter accessions, reflecting their phenotypic diversity and adaptability (Fig. 4). Spring accessions consistently exhibited a broader range of phenotypic variation across traits. For instance, HGW in spring accessions ranged from 6.72 to 43.75 g, representing a fourfold increase compared to the range observed in winter accessions, underscoring its potential for yield improvement. Moreover, spring accessions reached a Prot up to 29.83%, making them ideal for quality-focused breeding efforts.

Winter accessions, on the other hand, exhibited strong winterhardiness and capitalized on a longer flowering window, highlighting their adaptability to adverse conditions. Notably, PWSurv showed substantial variability, ranging from 30.34 to 79.27%, emphasizing the importance to prioritize selection for winterhardiness in pea breeding programmes (online resource 8). The limited phenotypic variation observed in winter populations, except for PWSurv, likely reflects both their smaller population size and the selective constraints imposed by winter conditions. Overall, these findings underscore the complementary potential of spring and winter accessions for breeding programmes. The contrasting traits between spring and winter groups highlight opportunities for targeted breeding focus, developing varieties tailored to different environmental conditions and consumer demands.

Convarieties demonstrate distinct performances across various traits

In the spring population, the analysed traits revealed varying levels of differences within and across convarieties (Fig. 5). Restrained variation in FTS among convarieties points to a consistent optimal flowering time window within the spring population. In contrast, plant height at flowering (PHFT) and HGW exhibited pronounced variability both within and across convarieties, with the widest ranges observed in the *sativum*, *speciosum* and unknown accessions. Prot also varied significantly, with *sativum* and *speciosum* convarieties achieving the highest mean values. The results consistently point to a clear association between convarietal classification and trait performance, despite unequal sample sizes across groups, emphasizing the relevance of taxonomic distinctions representing phenotypic differentiation within the spring pea population.

Convariety-dependent trait associations inform selection strategies within pea collection

Trait correlation analysis within the spring population revealed several significant associations, particularly among plant height and flowering-related traits, which

appeared tightly interconnected. In contrast, HGW, Spp (number of seeds per pod), and Prot showed weaker correlations with other traits, indicating more independent performances (Fig. 6). Inspection of correlations per convarieties highlighted distinct trends, suggesting different genetic mechanisms underlying the observed trait associations. For instance, in *axiphium*, Prot showed positive links with Ss (seed size) and HGW, but negative associations with FTdur (duration of flowering), hinting at the potential to combine large seeds, high protein, and a concentrated flowering period in this convariety. In contrast, *medullare* and *sativum* exhibited positive correlations between Prot and traits related to vegetative growth and earliness, while *speciosum* displayed a trade-off between seed size/weight and Spp (Online Resource 9). In the winter population, PWSurv showed a significant positive association with PHmax (maximum plant height), while other trends, including a potential trade-off with HGW, were less evident, likely reflecting the limitations imposed by small sample size (Online Resource 10). Moreover, significant association between plant height and flowering traits further reflected coordinated developmental dynamics between these traits. These findings underscore the genetic complexity governing the interplay of these important agronomic characteristics in pea, identifying the variation and the potential within the IPK pea populations to capitalize on by breeders.

Discussion

Unlocking the valuable, yet largely untapped, diversity within genetic resources is crucial for breeding efforts targeting the challenges posed by increasingly harsh environmental conditions. This need sparked an intensified interest in the curation and consolidation of historical phenotypic data to drive the transition from genebanks into bio-digital resource centres (Mascher et al. 2019). Building on the success of these previous efforts that focused primarily on wheat (Philipp et al. 2018; Svoboda et al. 2024) and barley (El Hanafi et al. 2023; González et al. 2018), we were motivated to unlock historical phenotypic data collected during regeneration for the crop species pea, using the IPK Gatersleben collection as a case study, which is the largest in the EU (Smýkal et al. 2012).

Selective outlier control enables robust insights from historical data

In line with previous studies (Berkner et al. 2024b; Bernal-Vasquez et al. 2016; El Hanafi et al. 2023; González et al. 2018; Philipp et al. 2018; Svoboda et al. 2024), our findings demonstrated that rigorous plausibility checks and

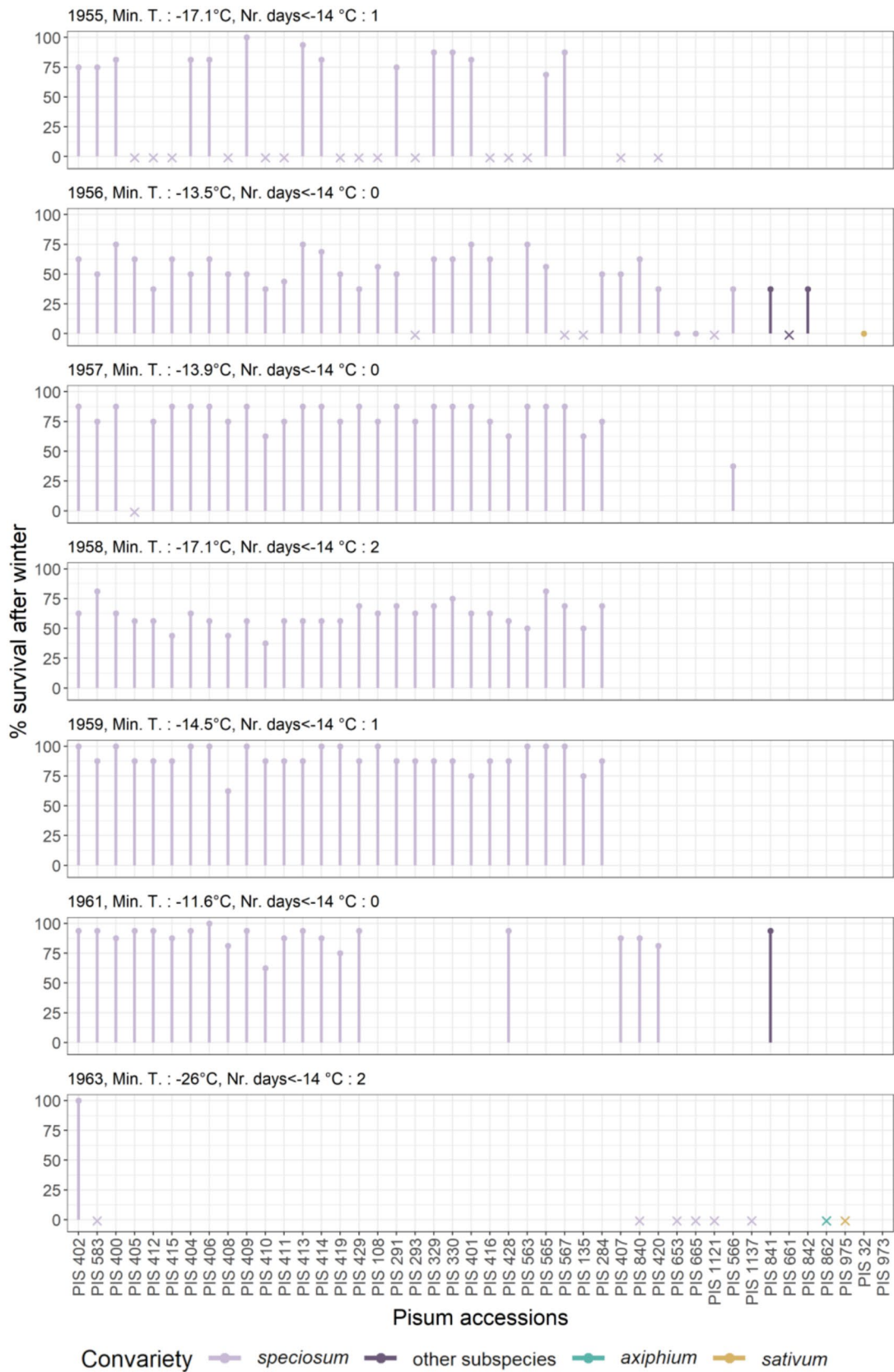


Fig. 3 Percentage post-winter survival (PWSurv) of each winter pea accession across 8 years. Crosses indicate years in which an accession was phenotyped for at least one post-winter trait, but PWSurv was not recorded. The yearly lowest minimum temperature at 5 cm above ground and number of days at which this temperature reached below -14°C are shown above each graph

systematic outlier removal substantially enhanced data quality as reflected by the increased heritability estimates across traits. Dependence on historical non-orthogonal data entails the risk of inaccurate extrapolation to contemporary conditions, especially when genotype-by-environment interactions are demonstrated (Dawson et al. 2013; He et al. 2016). However, in wheat high correlations were found between BLUEs obtained from historical and validation data (Philipp et al. 2018). We addressed the inherent challenge of non-orthogonality by employing linear mixed models that anchored year effects on frequently phenotyped accessions. This improved the accuracy of trait estimates for sparsely tested entries increasing the number of accessions with reliable phenotypic information (González et al. 2018; Philipp et al. 2018). In contrast to normalized rank products (Keilwagen et al. 2014), which rescale values to within-year ranks and omit true variance partitioning, and generic robust estimators (Morato and Stojanović 2021) that uniformly down-weight all outliers our pipeline combines trait-specific plausibility thresholds with mixed-model residual diagnostics. This hybrid strategy maintains proper extreme observations, which is critical to capturing seasonal or experimental outliers, while firmly restraining residual variance under the unreplicated and unbalanced data. Looking forward, this framework can be extended with cross-validation to fine-tune threshold settings or with hierarchical Bayesian outlier models that estimate error probabilities at observation level. In such way, we shift the phenotypic curation from a one-time task into a continuous data-driven calibration that steadily enhances heritability estimates and sets a new standard for genebank phenomics.

The diverse collection of IPK pea accessions inspires targeted breeding for various ideotypes

A diverse germplasm base is essential for unravelling complex trait architecture and driving breeding innovation. The IPK pea collection spans accessions from hot desert to subarctic climates (Köppen-Geiger classification; Peel et al. 2007), reflecting a wide ecological amplitude with potential implications for agronomic trait expression. The curated phenotypic data revealed that IPK retains exceptionally broad variation in key agronomic traits, demonstrating six-fold differences in grain weight, a 56 day span in

flowering time, and over 29.8% protein content among genebank accessions that go well beyond simple geographic or taxonomic structuring. For breeding purposes, this diversity can offer a powerful toolkit to match ideotypes adapted to varied climate requirements and markets demands (Carlson-Nilsson et al. 2021; Karkanis et al. 2016; Kuzbakova et al. 2022; Rubiales et al. 2021b). For instance, earliness confers dual benefits by mitigating terminal drought stress and reducing seed loss due to pod shattering, making it a key target for ideotype design in both temperate and marginal environments (Anderson 1955; Funatsuki et al. 2014; Lush et al. 1980; Parker et al. 2021). Moreover, improving protein content in pea is one of the central breeding objectives (Krefting 2017; Shanthakumar et al. 2022; Tao et al. 2017; Yang et al. 2023), to power its role in the protein transition as a key alternative to soybean in temperate regions (Krefting 2017; Shanthakumar et al. 2022). Simultaneously, combining protein content above 25% (Daba and Morris 2022; Yang et al. 2023) with hundred-seed weight over 32 g defines a high-value trait package that breeders can leverage for developing pea varieties with quality-driven breeding targets (Bundessortenamt 2024; Marton Genetics 2023; Producers and Growers Research Organisation 2025).

The distinct trait correlations observed within each variety underscore the value of moving beyond global core-sets to variety-specific selection indices (Chung and Liao 2022; Maranna et al. 2021; Najafi et al. 2024; Ouattara et al. 2024). Reframing ideotype performance as selection indices allows design of ideotype-specific core collections (Olivoto and Nardino 2020). On trait level, tailored core collections demonstrated increased statistical power, although the trade-off in increased population structure has to be managed (Berkner et al. 2024a). By maximizing diversity for the trait set shaping the ideotype, investigation of mechanisms defining crop suitability for key roles can be accelerated (Ravasi et al. 2020; Rocha et al. 2018). To ensure these sub-collections also perform in target environments, they can be further refined with the focused identification of germplasm strategy (FIGS, Khazaei et al. 2013), which uses environmental metadata to select accessions most likely to express the desired trait combinations under specific agroecological conditions. This two-step framework yields compact, high-resolution germplasm sets ideally suited for efficient pre-breeding.

Promising frost tolerance in IPK winter pea collection

The transition from spring to winter pea cultivation is key to mitigate spring and summer drought stress, an increasingly pressing issue under climate change, by permitting early

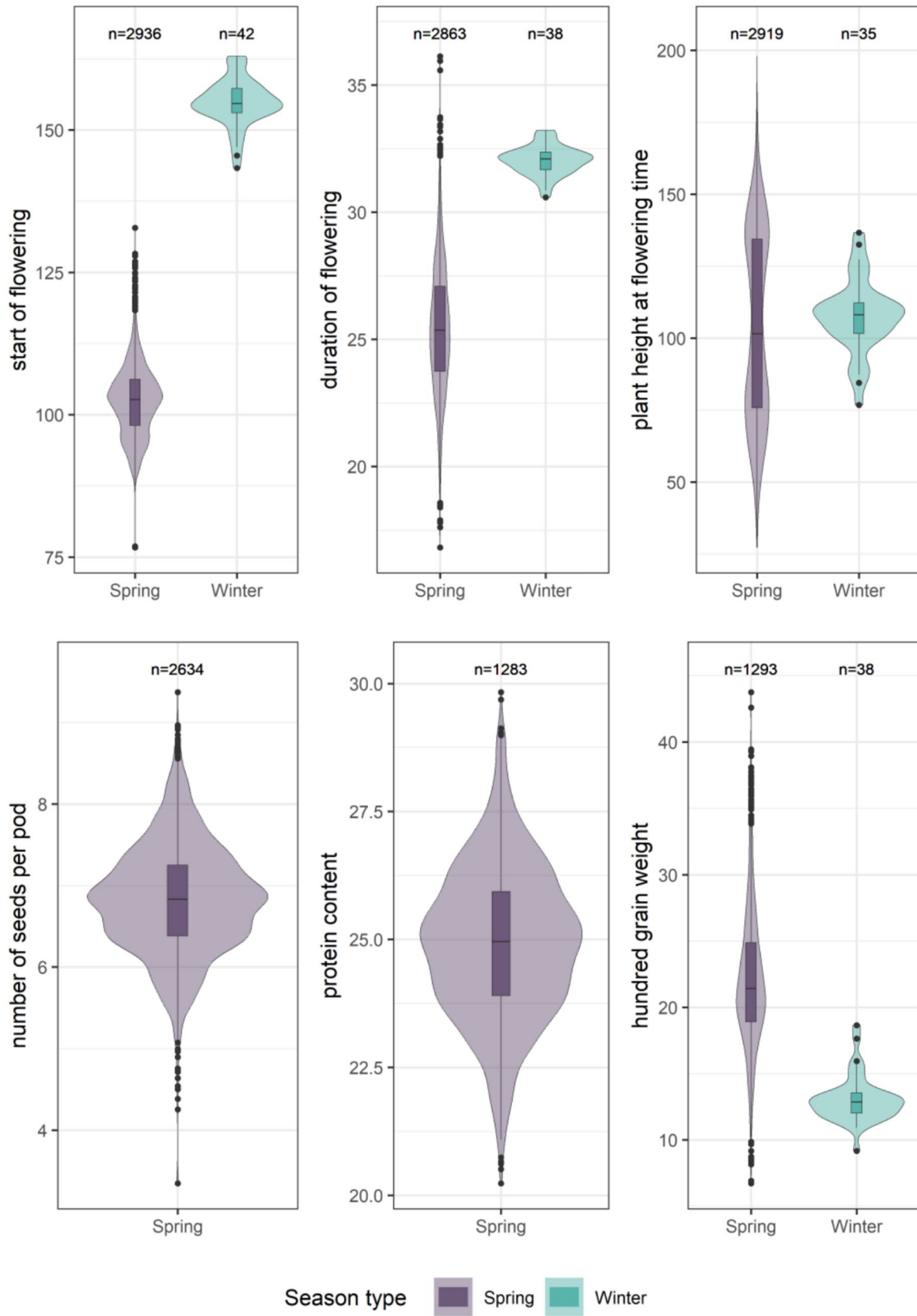


Fig. 4 Distribution of BLUEs for the traits start of flowering (days), duration of flowering (days), plant height (cm), number of seeds per pod, protein content (%), and hundred grain weight (g) for spring and winter accessions. The number of BLUEs is indicated above the corresponding graphs

root development and maturity. Furthermore, winter peas also offer notable advantages, including increased yield and enhanced soil benefits (Bagheri et al. 2023; Neugschwandtner et al. 2020; Urbatzka et al. 2012). However, the expansion of winter pea cultivation depends on the availability of consistently winter-hardy material (Homer et al. 2016; Karkanis et al. 2016). Success requires avoiding winter-kill while ensuring timely and vigorous spring growth. The historical winter data showed that *speciosum* whose flowering begins latest among winter types generally also achieved the highest post-winter survival (> 79%, $r=0.55$), whereas lines that flower earlier average below 50% survival. This suggests that delaying reproductive onset until after peak cold periods shields meristems from lethal freezes (Klein et al. 2014; Maqbool et al. 2010). Furthermore, the acceleration of growth and development preceding reproductive onset is associated to irreversible deacclimation, increasing the importance of strict regulation of these developmental progresses via, e.g. photoperiod sensitivity (Leinonen et al. 1997; Lejeune-Hénaut et al. 2008). These phenology/survival dynamics exhibit a Heritability of 0.84, indicating strong genetic control and reliable repeatability across divergent frost regimes.

Records of the IPK winter pea population cultivated in spring without adverse phenotypical effects indicate their possibly facultative identity, which would enable short breeding cycles and winter cultivation in regions with unreliable frost (Bauerle 2019). Moreover, occasional overwintering rates above 100% where new seedlings emerge during winter have been documented in peas and other crops and are attributed to cycling between primary physical dormancy and environmentally induced secondary dormancy (Batlla and Benech-Arnold 2006; Short et al. 2003; Urbatzka et al. 2011). In our data, unfavourable moisture and temperature fluctuations may similarly delay or re-impose dormancy, effectively postponing emergence until conditions become safe. While we lack direct dormancy measures, this dormancy–phenology coupling likely underpins the observed survival advantages and offers an additional, underexploited axis for breeding truly winter-hardy pea cultivars.

Towards an integrative approach for pea breeding

Recent EU Common Agricultural Policy and Rural Development Plan subsidies aim to expand legume acreage and

meet the European Green Deal’s “Farm to Fork” goals, yet pulses, including pea, still lag behind cereals due to lower yields, unstable production, and underinvestment in R&D (Magrini et al. 2016; Murphy-Bokern 2022; Rubiales et al. 2021a). Therefore, breeding pea must embrace integrative strategies that capitalize on both data-driven insights and cutting-edge technology. Historical data reveal phenotypic plasticity across decades and diverse climates. Leveraging this information enables the identification of accessions adapted to specific agroecological niches, guiding the design of next-generation ideotypes (Alseekh et al. 2025). Yet full-scale phenotyping remains costly and time-intensive, leaving substantial portions of the collection only sparsely characterized. Cost-effective high-throughput genotyping offers a pragmatic workaround that can safeguard genetic integrity by flagging allele-frequency shifts that signal drift during regeneration and furnish the dense marker panels needed for downstream analyses. Although the large and repetitive pea genome long impeded genomic research, targeted core collections from USDA and European genebanks have nonetheless enabled GWAS to uncover major-effect loci albeit with limited mapping resolution due to sparse SNP coverage (Burstin et al. 2007; Crosta et al. 2023; Irzykowska and Wolko 2004; Klein et al. 2020). In parallel, genomic prediction has emerged as a powerful extension of genome wide-association studies (GWAS), capable of aggregating small-effect loci to forecast complex traits (Goddard 2009; Meuwissen et al. 2001). However, the effectiveness of these predictions can be compromised when applied to populations with limited genetic diversity or when environmental conditions differ significantly from those encountered during model training (Wientjes et al. 2015). This highlights the need for ongoing refinement of prediction methodologies to ensure robust predictions across diverse breeding scenarios (Bari et al. 2021; Crosta et al. 2023; Tayeh et al. 2015). Across-genebank predictions can address this limitation by training on a broader spectrum of allelic diversity and environmental contexts. For instance, models built on one repository still achieved moderate predictive power when tested in another, despite differences in allele frequencies and trial protocols (El Hanafi et al. 2023; Schulthess et al. 2022). Therefore, unlocking the phenotype data of the IPK genebank collection, anticipating a future genotyping endeavour similar to the USDA and JIC UK (Cheng et al. 2015; Jing et al. 2010), is valuable not only for the collection itself, but once combined with genotype data, also provides a resource to characterize other collections. By pooling marker data, trial metadata, and phenotypes across multiple collections, genebank curators can mitigate both genetic-base and environment-shift limitations, allowing them to impute unphenotyped traits (e.g. frost tolerance, disease resistance)

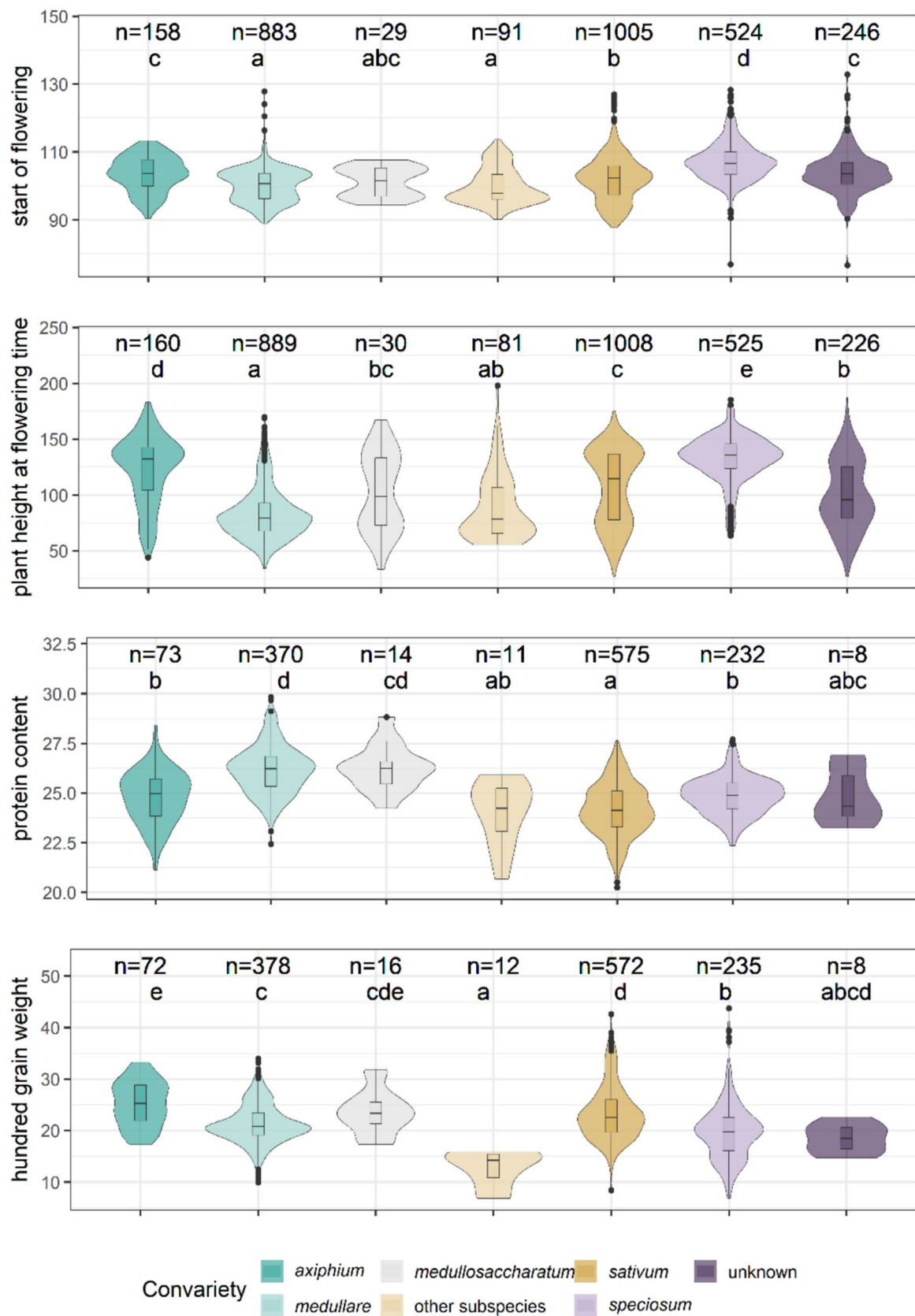


Fig. 5 Distribution of BLUEs for start of flowering (days), plant height (cm), protein content (%), and hundred grain weight (g) across cultivars in the spring population. The numbers of BLUEs are given above the plots along with the compact letter display indicating the significance groups. The absence of overlapping letters indicate significant differences ($p < 0.05$) between cultivars

in sparsely characterized accessions without additional field trials. Integrating such across-genebank predictions with machine-learning algorithms that prioritize haplotype-trait associations further enables creation of compact, high-resolution training sets tailored to specific breeding objectives.

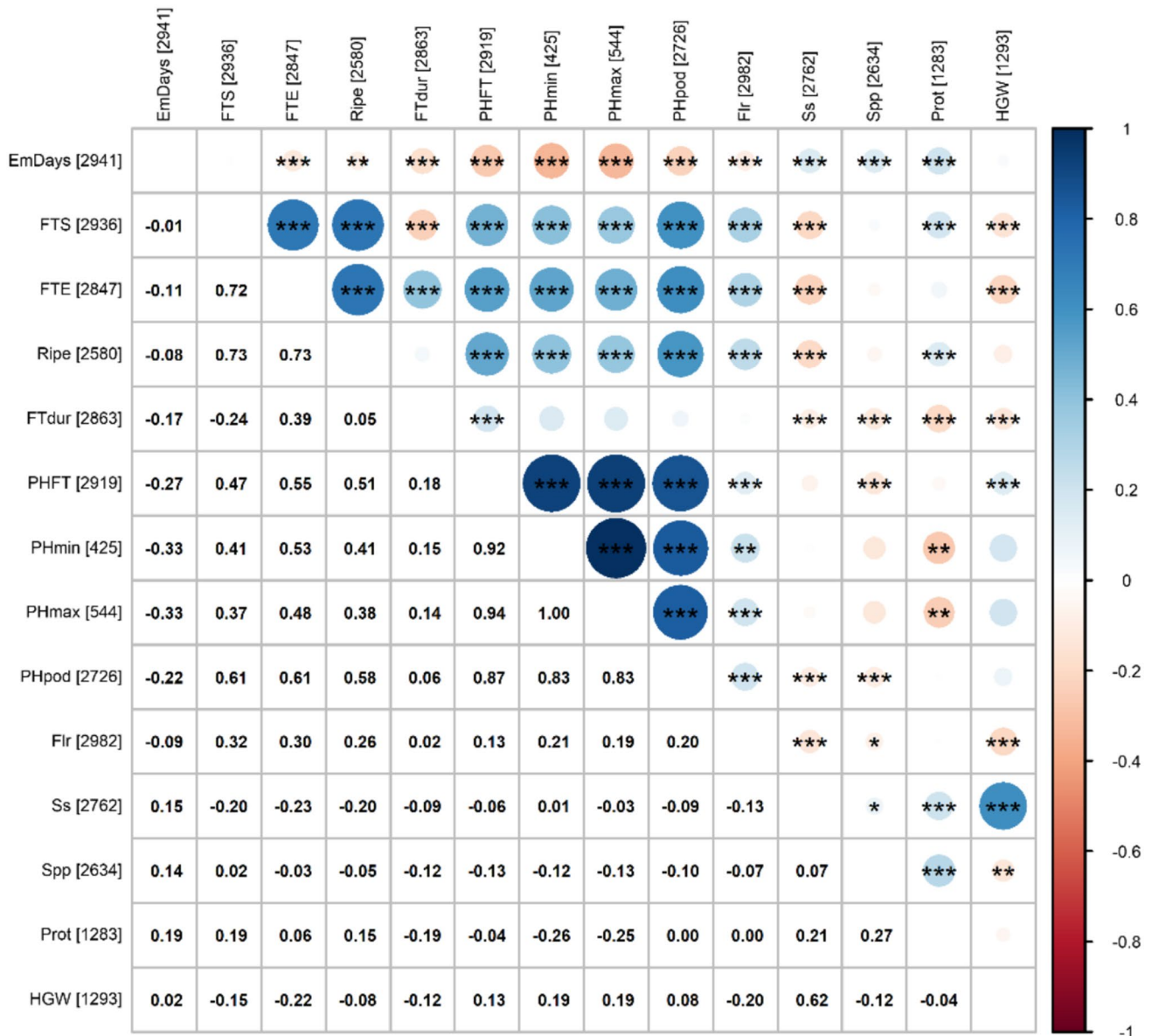


Fig. 6 Trait correlations in the spring pea population based on the estimated BLUEs. Spearman's correlation coefficients between traits are shown below the diagonal; corresponding visualizations using coloured symbols are shown above. Numbers in parentheses indicate the number of BLUEs available for each trait. Asterisks denote significance after Holm's multiple testing correction ($p < 0.05$, $p < 0.01$, $p < 0.001$). Colour and size of the circles reflect the strength and direction of the correlation, with blue indicating positive and red negative associations. EmDays=days to emergence

(days since March 1th), FTS=start of flowering (days since March 1th), FTE=end of flowering (days since March 1th), Ripe=date of picking ripeness (days since March 1th), FTdur=duration of flowering (days since March 1th), PHFT=plant height at flowering time (cm), PHmin=minimum plant height (cm), PHmax=maximum plant height (cm), PHpod=height of the first pod (cm), Flr=number of flowers per inflorescence, Ss=seed size, Spp=number of seeds per pod, Prot=protein content (%), HGW=hundred grain weight (g)

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-025-05032-5>.

Acknowledgements This publication had never been possible without the dedication of the former and current employees of The Federal Ex Situ Genebank for Agricultural and Horticultural Crops, who recorded and preserved the extensive phenotypic data.

Author contributions statement SEH and JCR provided the key ideas of the review and the methodology. SEH, JCR, and LGO provided the project administration. MK and UL performed the investigation. Data compilation and management were performed by MK, UL, MO, and SW. LN and SEH enacted the computational analysis and created the visualization. Supervision was provided by SEH. Data validation was performed by LN, SEH, and VM. Funding was enabled by SEH, VM, LGO, and JCR. The initial draft was compiled by LN, SEH, and JCR. All authors contributed to the final reading and approved the submitted version.

Funding Open Access funding enabled and organized by Projekt DEAL. We acknowledge the financial support provided by the German Federal Ministry of Education and Research under the frame of the DiPisum-Project (Grant No. 031B1444A), of the European Union through Horizon Europe under the frame of the Legume Generation Project (Grant No. 101081329), and of the State of Saxony-Anhalt (Germany) under the project “Unleashing Performance and Sustainability through Genotypes and Crop Mixtures” (UpMix, LSA I 228).

Data availability A data publication will become available shortly. In the meantime, data can be requested from the corresponding author.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

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/>.

References

- Alefeld F (1866) Landwirtschaftliche Flora, Berlin
- Aalsekh S, Klemmer A, Yan J, Guo T, Fernie AR (2025) Embracing plant plasticity or robustness as a means of ensuring food security. *Nat Commun* 16:461. <https://doi.org/10.1038/s41467-025-55872-4>
- Anderson SR (1955) Development of pods and seeds of birdsfoot trefoil, *Lotus corniculatus* L., as related to maturity and to seed yields. *Agron J* 47:483–487. <https://doi.org/10.2134/agronj1955.00021962004700100009x>
- Anglin NL, Amri A, Kehel Z, Ellis D (2018) A case of need: linking traits to genebank accessions. *Biopreserv Biobank* 16:337–349. <https://doi.org/10.1089/bio.2018.0033>
- Annicchiarico P (2008) Adaptation of cool-season grain legume species across climatically-contrasting environments of southern Europe. *Agron J* 100:1647–1654. <https://doi.org/10.2134/agronj2008.0085>
- Annicchiarico P, Romani M, Cabassi G, Ferrari B (2017) Diversity in a pea (*Pisum sativum*) world collection for key agronomic traits in a rain-fed environment of southern Europe. *Euphytica* 213:245. <https://doi.org/10.1007/s10681-017-2033-y>
- Bagheri M, Santos CS, Rubiales D, Vasconcelos MW (2023) Challenges in pea breeding for tolerance to drought: status and prospects. *Ann Appl Biol* 183:108–120. <https://doi.org/10.1111/aab.12840>
- Bari MAA, Zheng P, Viera I, Worrall H, Szwiec S, Ma Y, Main D, Coyne CJ et al (2021) Harnessing genetic diversity in the USDA pea germplasm collection through genomic prediction. *Front Genet*. <https://doi.org/10.3389/fgene.2021.707754>
- Batlla D, Benech-Arnold RL (2006) The role of fluctuations in soil water content on the regulation of dormancy changes in buried seeds of *Polygonum aviculare* L. *Seed Sci Res* 16:47–59. <https://doi.org/10.1079/SSR2005234>
- Bauerle WL (2019) Disentangling photoperiod from hop vernalization and dormancy for global production and speed breeding. *Sci Rep* 9:16003. <https://doi.org/10.1038/s41598-019-52548-0>
- Berkner MO, Jiang Y, Reif JC, Schulthess AW (2024a) Trait-customized sampling of core collections from a winter wheat genebank collection supports association studies. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2024.1451749>
- Berkner MO, Weise S, Reif JC, Schulthess AW (2024b) Genomic prediction reveals unexplored variation in grain protein and lysine content across a vast winter wheat genebank collection. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2023.1270298>
- Bernal-Vasquez A-M, Utz H-F, Piepho H-P (2016) Outlier detection methods for generalized lattices: a case study on the transition from ANOVA to REML. *Theor Appl Genet* 129:787–804. <https://doi.org/10.1007/s00122-016-2666-6>
- Bundessortenamt (2024) Beschreibende sortenliste getreide, Mais Öl-und faserpflanzen, Leguminosen, Rüben, Zwischenfrüchte. https://www.bundessortenamt.de/bsa/media/Files/BSL/bsl_getreide_2024.pdf
- Burstin J, Marget P, Huart M, Moessner A, Mangin B, Duchene C, Desprez B, Munier-Jolain N et al (2007) Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. *Plant Physiol* 144:768–781. <https://doi.org/10.1104/pp.107.096966>
- Butler D, Cullis B, Gilmour A, Gogel B, Thompson R (2017) ASReml-R reference manual version 4. VSN International Ltd, UK
- Carlson-Nilsson U, Aloisi K, Vågen IM, Rajala A, Møllmann JB, Rasmussen SK, Niemi M, Wojciechowska E et al (2021) Trait expression and environmental responses of pea (*Pisum sativum* L.) genetic resources targeting cultivation in the Arctic. *Front Plant Sci* 12:688067. <https://doi.org/10.3389/fpls.2021.688067>
- Carrouée B, Crépon K, Peyronnet C (2003) Les protéagineux: intérêt dans les systèmes de production fourragers français et européens. *Fourrages*. 174:163–182
- Cheng P, Holdsworth W, Ma Y, Coyne CJ, Mazourek M, Grusak MA, Fuchs S, McGee RJ (2015) Association mapping of agronomic and quality traits in USDA pea single-plant collection. *Mol Breed* 35:75. <https://doi.org/10.1007/s11032-015-0277-6>
- Chung P-Y, Liao C-T (2022) Selection of parental lines for plant breeding via genomic prediction. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2022.934767>

- Crosta M, Romani M, Nazzicari N, Ferrari B, Annicchiarico P (2023) Genomic prediction and allele mining of agronomic and morphological traits in pea (*Pisum sativum*) germplasm collections. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2023.1320506>
- Daba SD, Morris CF (2022) Pea proteins: variation, composition, genetics, and functional properties. *Cereal Chem* 99:8–20. <https://doi.org/10.1002/cche.10439>
- Dawson JC, Endelman JB, Heslot N, Crossa J, Poland J, Dreisigacker S, Manès Y, Sorrells ME et al (2013) The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Res* 154:12–22. <https://doi.org/10.1016/j.fcr.2013.07.020>
- Dempewolf H, Baute G, Anderson J, Kilian B, Smith C, Guarino L (2017) Past and future use of wild relatives in crop breeding. *Crop Sci* 57:1070–1082. <https://doi.org/10.2135/cropsci2016.10.0885>
- El Hanafi S, Jiang Y, Kehel Z, Schulthess AW, Zhao Y, Mascher M, Haupt M, Himmelbach A et al (2023) Genomic predictions to leverage phenotypic data across genebanks. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2023.1227656>
- Engels JMM, Ebert AW, van Hintum T (2024) Collaboration between private and public genebanks in conserving and using plant genetic resources. *Plants* 13:247. <https://doi.org/10.3390/plants13020247>
- FAO (2024) Crops and livestock products. <https://www.fao.org/faostat/en/#data/QCL>
- Franguelli N, Cavalli D, Notario T, Pecetti L, Annicchiarico P (2024) Frost tolerance improvement in pea and white lupin by a high-throughput phenotyping platform. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2024.1490577>
- Funatsuki H, Suzuki M, Hirose A, Inaba H, Yamada T, Hajika M, Komatsu K, Katayama T et al (2014) Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc Natl Acad Sci USA* 111:17797–17802. <https://doi.org/10.1073/pnas.1417282111>
- Marton Genetics (2023) Seed catalogue 2022–2023 (ENG). <https://martongenetics.com/en/catalogues/>
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257. <https://doi.org/10.1007/s10709-008-9308-0>
- González MY, Philipp N, Schulthess AW, Weise S, Zhao Y, Börner A, Oppermann M, Graner A et al (2018) Unlocking historical phenotypic data from an ex situ collection to enhance the informed utilization of genetic resources of barley (*Hordeum* sp.). *Theor Appl Genet* 131:2009–2019. <https://doi.org/10.1007/s00122-018-3129-z>
- Green FN (2008) Classification of *Pisum sativum* subsp *sativum* cultivars into groups and subgroups using simply inherited characters. 799 edn International Society for Horticultural Science (ISHS), Belgium, pp 155–162. <https://doi.org/10.17660/ActaHortic.2008.799.22>
- Guerra D, Morcia C, Badeck F, Rizza F, Delbono S, Francia E, Milc JA, Monostori I et al (2022) Extensive allele mining discovers novel genetic diversity in the loci controlling frost tolerance in barley. *Theor Appl Genet* 135:553–569. <https://doi.org/10.1007/s00122-021-03985-x>
- He S, Schulthess AW, Mirdita V, Zhao Y, Korzun V, Bothe R, Ebmeyer E, Reif JC et al (2016) Genomic selection in a commercial winter wheat population. *Theor Appl Genet* 129:641–651. <https://doi.org/10.1007/s00122-015-2655-1>
- He S, Liu H, Zhan J, Meng Y, Wang Y, Wang F, Ye G (2022) Genomic prediction using composite training sets is an effective method for exploiting germplasm conserved in rice gene banks. *Crop J* 10:1073–1082. <https://doi.org/10.1016/j.cj.2021.11.011>
- Holland JB, Nyquist WE, Cervantes-Martínez CT, Janick J (2003) Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev* 22:9–122. <https://doi.org/10.1002/9780470650202.ch2>
- Homer A, ŞahİN M, Küçüközdemir Ü (2016) Evaluation of pea (*Pisum sativum* L.) germplasm for winter hardiness in Central Anatolia, Turkey, using field and controlled environment. <https://doi.org/10.17221/186/2015-CJGPB>
- Irzykowska L, Wolko B (2004) Interval mapping of QTLs controlling yield-related traits and seed protein content in *Pisum sativum*. *J Appl Genet* 45:297–306
- Jiang Y, Weise S, Graner A, Reif JC (2021) Using genome-wide predictions to assess the phenotypic variation of a barley (*Hordeum* sp.) gene bank collection for important agronomic traits and passport information. *Front Plant Sci* 11:604781. <https://doi.org/10.3389/fpls.2020.604781>
- Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D, Ambrose MJ, Ellis THN et al (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evol Biol* 10:44. <https://doi.org/10.1186/1471-2148-10-44>
- Karkanis A, Ntatsi G, Kontopoulou C-K, Pristeri A, Bilalis D, Savvas D (2016) Field pea in European cropping systems: adaptability, biological nitrogen fixation and cultivation practices. *Not Bot Horti Agrobot Cluj-Napoca* 44:325–336. <https://doi.org/10.15835/nbha44210618>
- Keilwagen J, Kilian B, Özkan H, Babben S, Perovic D, Mayer KF, Walther A, Poskar CH et al (2014) Separating the wheat from the chaff: a strategy to utilize plant genetic resources from ex situ genebanks. *Sci Rep* 4:5231. <https://doi.org/10.1038/srep05231>
- Khazaei H, Street K, Bari A, Mackay M, Stoddard FL (2013) The FIGS (focused identification of germplasm strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLoS One* 8:e63107. <https://doi.org/10.1371/journal.pone.0063107>
- Klein A, Houtin H, Rond C, Marget P, Jacquin F, Boucherot K, Huart M, Rivière N et al (2014) QTL analysis of frost damage in pea suggests different mechanisms involved in frost tolerance. *Theor Appl Genet* 127:1319–1330. <https://doi.org/10.1007/s00122-014-2299-6>
- Klein A, Houtin H, Rond-Coissieux C, Naudet-Huart M, Touratier M, Marget P, Burstin J (2020) Meta-analysis of QTL reveals the genetic control of yield-related traits and seed protein content in pea. *Sci Rep* 10:15925. <https://doi.org/10.1038/s41598-020-72548-9>
- Krefting J (2017) The appeal of pea protein. *J Ren Nutr* 27:e31–e33. <https://doi.org/10.1053/j.jrn.2017.06.009>
- Kuzbakova M, Khassanova G, Oshergina I, Ten E, Jatayev S, Yerzhebayeva R, Bulatova K, Khalbayeva S et al (2022) Height to first pod: a review of genetic and breeding approaches to improve combine harvesting in legume crops. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2022.948099>
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. <https://doi.org/10.1186/1471-2105-9-559>
- Langfelder P, Horvath S (2012) Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw* 46:1–17. <https://doi.org/10.18637/jss.v046.i11>
- Lehmann CO (1954) Das morphologische system der saaterbsen. *Der Züchter* 24:316–337. <https://doi.org/10.1007/BF00710392>
- Leinonen I, Repo T, Hänninen H (1997) Changing environmental effects on frost hardiness of Scots pine during dehardening. *Ann Bot* 79:133–138. <https://doi.org/10.1006/anbo.1996.0321>

- Lejeune-Hénaut I, Hanocq E, Béthencourt L, Fontaine V, Delbreil B, Morin J, Petit A, Devaux R et al (2008) The flowering locus *Hr* colocalizes with a major QTL affecting winter frost tolerance in *Pisum sativum* L. *Theor Appl Genet* 116:1105–1116. <https://doi.org/10.1007/s00122-008-0739-x>
- Lenth RV (2024) Emmeans: estimated marginal means, aka least-squares means. R package version 1.10.4. <https://CRAN.R-project.org/package=emmeans>
- Liu R, Fang L, Yang T, Zhang X, Hu J, Zhang H, Han W, Hua Z et al (2017) Marker-trait association analysis of frost tolerance of 672 worldwide pea (*Pisum sativum* L.) collections. *Sci Rep* 7:5919. <https://doi.org/10.1038/s41598-017-06222-y>
- Lush WM, Evans LT, Wien HC (1980) Environmental adaptation of wild and domesticated cowpeas (*Vigna unguiculata* (L.) walp.). *Field Crops Res* 3:173–187. [https://doi.org/10.1016/0378-4290\(80\)90023-4](https://doi.org/10.1016/0378-4290(80)90023-4)
- Magrini M-B, Anton M, Cholez C, Corre-Hellou G, Duc G, Jeuffroy M-H, Meynard J-M, Pelzer E et al (2016) Why are grain-legumes rarely present in cropping systems despite their environmental and nutritional benefits? Analyzing lock-in in the French agrifood system. *Ecol Econ* 126:152–162. <https://doi.org/10.1016/j.ecolecon.2016.03.024>
- Maqbool A, Shafiq S, Lake L (2010) Radiant frost tolerance in pulse crops: a review. *Euphytica* 172:1–12. <https://doi.org/10.1007/s10681-009-0031-4>
- Maranna S, Nataraj V, Kumawat G, Chandra S, Rajesh V, Ramteke R, Patel RM, Ratnaparkhe MB et al (2021) Breeding for higher yield, early maturity, wider adaptability and waterlogging tolerance in soybean (*Glycine max* L.): a case study. *Sci Rep*. <https://doi.org/10.1038/s41598-021-02064-x>
- Mascher M, Schreiber M, Scholz U, Graner A, Reif JC, Stein N (2019) Genebank genomics bridges the gap between the conservation of crop diversity and plant breeding. *Nat Genet* 51:1076–1081. <https://doi.org/10.1038/s41588-019-0443-6>
- McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D et al (2013) Feeding the future. *Nature* 499:23–24. <https://doi.org/10.1038/499023a>
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>
- Milner SG, Jost M, Taketa S, Mazón ER, Himmelbach A, Oppermann M, Weise S, Knüpffer H et al (2019) Genebank genomics highlights the diversity of a global barley collection. *Nat Genet* 51:319–326. <https://doi.org/10.1038/s41588-018-0266-x>
- Morato MM, Stojanović V (2021) A robust identification method for stochastic nonlinear parameter varying systems. *Math Modelling Control* 1:35–51. <https://doi.org/10.3934/mmc.2021004>
- Moritz S, Bartz-Beielstein T (2017) imputeTS: time series missing value imputation in R. *The R Journal* 9:207–218
- Müntz K, Lehmann CO (1987) Reserveproteinforschung und Genbank. *Die Kulturpflanze* 35:25–52. <https://doi.org/10.1007/BF02163328>
- Murphy-Bokern D (2022) Developing legume-supported cropping systems in Europe: Have we overlooked something? *Ann Appl Biol* 181:133–136. <https://doi.org/10.1111/aab.12764>
- Najafi S, Shoostari L, Mohammadi R, Etminan A, Mehrabi A-M, Ghorbanpour M (2024) Genetic diversity and population structure analysis in durum wheat (*Triticum turgidum* L. var. *Durum*) germplasm using agronomic characters and CBDP markers. *Genet Resour Crop Evol* 71:3405–3417. <https://doi.org/10.1007/s10722-023-01854-8>
- Neuschwandtner RW, Bernhuber A, Kammlander S, Wagentristl H, Klimek-Kopyra A, Kaul HP (2020) Yield structure components of autumn- and spring-sown pea (*Pisum sativum* L.). *Acta Agriculturae Scandinavica. Sect b: Soil Plant Sci* 70(2):109–116
- Olivoto T, Nardino M (2020) MGIDI: towards an effective multivariate selection in biological experiments. *Bioinformatics* 37:1383–1389. <https://doi.org/10.1093/bioinformatics/btaa981>
- Oppermann M, Weise S, Dittmann C, Knüpffer H (2015) GBIS: the information system of the German Genebank. *Database* 2015:bav21. <https://doi.org/10.1093/database/bav021>
- Osman SA, Ali H (2021) Genetic relationship of some *Pisum sativum* subspecies using different molecular markers. *Jordan J Biol Sci* 14:1. <https://doi.org/10.54319/jjbs/140109>
- Ouattara F, Agre PA, Adejumbi II, Akoroda MO, Sorho F, Ayolié K, Bhattacharjee R (2024) Multi-trait selection index for simultaneous selection of water yam (*Dioscorea alata* L.) genotypes. *Agronomy* 14:128. <https://doi.org/10.3390/agronomy14010128>
- Parker TA, Lo S, Gepts P (2021) Pod shattering in grain legumes: emerging genetic and environment-related patterns. *Plant Cell* 33:179–199. <https://doi.org/10.1093/plcell/koaa025>
- Peel MC, Finlayson BL, McMahon TA (2007) Updated world map of the Köppen–Geiger climate classification. *Hydrol Earth Syst Sci* 11:1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
- Philipp N, Weise S, Oppermann M, Börner A, Graner A, Keilwagen J, Kilian B, Zhao Y et al (2018) Leveraging the use of historical data gathered during seed regeneration of an *ex Situ* Genebank collection of wheat. *Front Plant Sci* 9:609. <https://doi.org/10.3389/fpls.2018.00609>
- Processors and Growers Research Organisation (2025) Combining peas: PGRO descriptive list 2025. <https://www.pgro.org/downloads/DL-LIST-PDF.pdf>
- Qiu L-J, Xing L-L, Guo Y, Wang J, Jackson SA, Chang R-Z (2013) A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Mol Biol* 83:41–50. <https://doi.org/10.1007/s11103-013-0076-6>
- R Core Team (2024) R: a language and environment for statistical computing. <https://www.R-project.org>
- Ravasi RA, Paleari L, Vesely FM, Movedi E, Thoeke W, Confalonieri R (2020) Ideotype definition to adapt legumes to climate change: a case study for field pea in Northern Italy. *Agric for Meteorol* 291:108081. <https://doi.org/10.1016/j.agrformet.2020.108081>
- Rocha JRdASdC, Machado JC, Carneiro PCS (2018) Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy* 10:52–60. <https://doi.org/10.1111/gcbb.12443>
- Rogers H, Dora M, Tsolakis N, Kumar M (2024) Plant-based food supply chains: recognising market opportunities and industry challenges of pea protein. *Appl Food Res* 4:100440. <https://doi.org/10.1016/j.afres.2024.100440>
- Roorkiwal M, Bharadwaj C, Barmukh R, Dixit GP, Thudi M, Gaur PM, Chaturvedi SK, Fikre A et al (2020) Integrating genomics for chickpea improvement: achievements and opportunities. *Theor Appl Genet* 133:1703–1720. <https://doi.org/10.1007/s00122-020-03584-2>
- Rubiales D, Annicchiarico P, Vaz Patto MC, Julier B (2021a) Legume breeding for the agroecological transition of global agri-food systems: a European perspective. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2021.782574>
- Rubiales D, Osuna-Caballero S, González-Bernal MJ, Cobos MJ, Flores F (2021b) Pea breeding lines adapted to autumn sowings in broomrape prone mediterranean environments. *Agronomy* 11:769. <https://doi.org/10.3390/agronomy11040769>
- Schulthess AW, Kale SM, Liu F, Zhao Y, Philipp N, Rembe M, Jiang Y, Beukert U et al (2022) Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement. *Nat Genet* 54:1544–1552. <https://doi.org/10.1038/s41588-022-01189-7>
- Shanthakumar P, Klepacka J, Bains A, Chawla P, Dhull SB, Najda A (2022) The Current situation of pea protein and its application in the food industry. *Molecules* 27:5354. <https://doi.org/10.3390/molecules27165354>

- Short RW, Stratton RG, Muehlbauer FJ, McPhee KE, Chen W (2003) Winter pea trial results. 2002 progress report. <http://pwa.ars.usda.gov/pullmann/glgp/>
- Smýkal P, Aubert G, Burstin J, Coyne CJ, Ellis NTH, Flavell AJ, Ford R, Hýbl M et al (2012) Pea (*Pisum sativum* L.) in the genomic era. *Agronomy* 2:74–115. <https://doi.org/10.3390/agronomy2020074>
- Svoboda P, Holubec V, Reif JC, Berkner MO (2024) Curation of historical phenotypic wheat data from the Czech Genebank for research and breeding. *Scientific Data* 11:763. <https://doi.org/10.1038/s41597-024-03598-1>
- Tao A, Afshar RK, Huang J, Mohammed YA, Espe M, Chen C (2017) Variation in yield, starch, and protein of dry pea grown across montana. *Agron J* 109:1491–1501. <https://doi.org/10.2134/agronj2016.07.0401>
- Tayeh N, Klein A, Le Paslier MC, Jacquin F, Houtin H, Rond C, Chabert-Martinello M, Magnin-Robert JB et al (2015) Genomic Prediction in pea: effect of marker density and training population size and composition on prediction accuracy. *Front Plant Sci* 6:941. <https://doi.org/10.3389/fpls.2015.00941>
- Thavarajah D, Lawrence T, Boatwright L, Windsor N, Johnson N, Kay J, Shipe E, Kumar S et al (2023) Organic dry pea (*Pisum sativum* L.): a sustainable alternative pulse-based protein for human health. *PLoS ONE* 18:e0284380. <https://doi.org/10.1371/journal.pone.0284380>
- Trnka M, Kocmánková E, Balek J, Eitzinger J, Ruget F, Formayer H, Hlavinka P, Schaumberger A et al (2010) Simple snow cover model for agrometeorological applications. *Agric for Meteorol* 150:1115–1127. <https://doi.org/10.1016/j.agrformet.2010.04.012>
- Urbatzka P, Graß R, Haase T, Schüler C, Trautz D, Heß J (2011) Grain yield and quality characteristics of different genotypes of winter pea in comparison to spring pea for organic farming in pure and mixed stands. *Org Agric* 1:187–202. <https://doi.org/10.1007/s13165-011-0015-2>
- Urbatzka P, Graß R, Haase T, Schüler C, Heß J (2012) Influence of different sowing dates of winter pea genotypes on winter hardiness and productivity as either winter catch crop or seed legume. *Eur J Agron* 40:112–119. <https://doi.org/10.1016/j.eja.2012.03.001>
- Vogt-Kaute W, Quendt U (2022) Wintererbse: Anbautelegramm. https://www.legunet.de/fileadmin/legunet/Dokumente/legunet_Website_Anbautelegramm_Wintererbse_220920.pdf
- Weise S, Blattner FR, Börner A, Dehmer KJ, Grube M, Harpke D, Lohwasser U, Oppermann M et al (2025) The German Federal Ex Situ Genebank for agricultural and horticultural crops: conservation, exploitation and steps towards a bio-digital resource centre. *Genet Resour* 2:91–105. <https://doi.org/10.46265/genesj.GYDY5145>
- Wientjes YCJ, Bijma P, Veerkamp RF, Calus MPL (2015) An equation to predict the accuracy of genomic values by combining data from multiple traits, populations, or environments. *Genetics* 202:799–823. <https://doi.org/10.1534/genetics.115.183269>
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W et al (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3:160018. <https://doi.org/10.1038/sdata.2016.18>
- Windsor N, Boatwright L, Boyles R, Bridges W, Rubiales D, Thavarajah D (2024) Characterizing dry pea (L) for improved nutritional traits and the potential for biofortification. *Legume Sci* 6:e250. <https://doi.org/10.1002/leg3.250>
- Yang J, Lorenzetti RL, Bing D, Zhang S, Lu J, Chen L (2023) Composition, functionalities, and digestibility of proteins from high protein and normal pea (*Pisum sativum*) genotypes. *Sustainable Food Proteins* 1:4–15. <https://doi.org/10.1002/sfp2.1005>
- Zhang X, Wan S, Hao J, Hu J, Yang T, Zong X (2016) Large-scale evaluation of pea (*Pisum sativum* L.) germplasm for cold tolerance in the field during winter in Qingdao. *The Crop J* 4:377–383. <https://doi.org/10.1016/j.cj.2016.06.016>

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