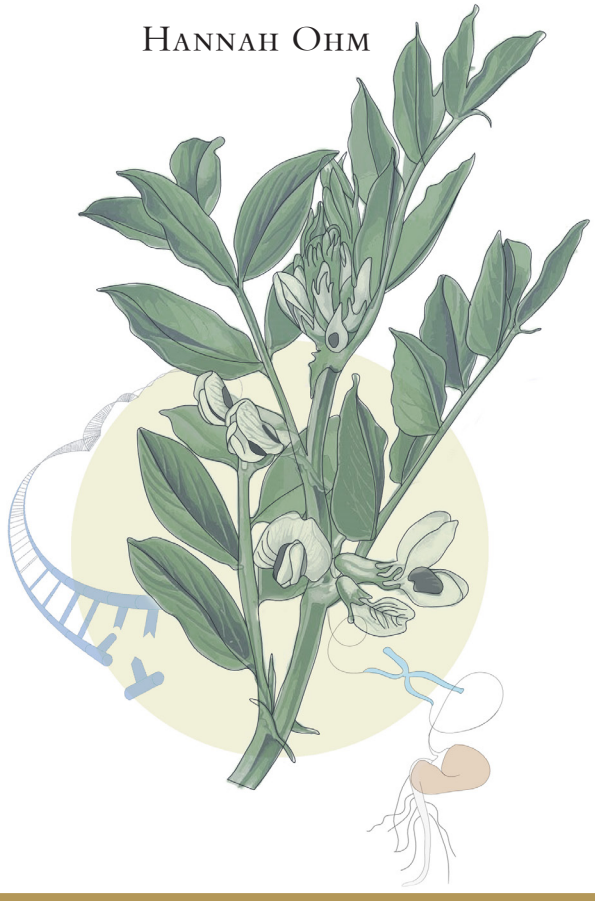
Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2025:20
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Genetic insights into seed development, flowering and diversity in faba bean

Pre-breeding for sustainable agriculture

Hannah Ohm



Genetic insights into seed development, flowering and diversity in faba bean

Pre-breeding for sustainable agriculture

Hannah Ohm
Faculty of Landscape Architecture, Horticulture and Crop Production Sciences
Department of Plant Breeding
Alnarp

SLU
SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

DOCTORAL THESIS
Alnarp 2025

Acta Universitatis Agriculturae Sueciae
2025:20

ISSN 1652-6880
ISBN (print version) 978-91-8046-455-0
ISBN (electronic version) 978-91-8046-505-2
https://doi.org/10.54612/a.7uf5cpud12

Swedish University of Agricultural Sciences, Department of Plant Breeding, Alnarp, Sweden
Print: SLU Grafisk service, Alnarp 2025

Genetic insights into seed development, flowering and diversity in faba bean

Abstract

Faba bean (*Vicia faba* L.) is a nutrient-rich legume with significant potential for sustainable agriculture in northern climates, yet breeding advancements remain limited. This thesis integrates phenotypic, genetic, and biochemical analyses to develop genomic tools and accelerate the improvement of key agronomic traits and seed quality. A global diversity panel of 220 faba bean accessions, evaluated across two growing seasons, revealed substantial phenotypic variation in flowering time, plant height, yield, and seed composition. Notably, no yield penalty was observed for higher protein content, challenging traditional trade-offs and suggesting potential for simultaneous improvement. A genome-wide association study (GWAS) identified 51 SNP markers linked to ten agronomic traits. Transcriptomic and biochemical analyses of seed development elucidated the temporal regulation of nutrient accumulation, revealing early protein deposition in both embryo and endosperm, followed by progressive starch accumulation and consistently low lipid content. Gene expression analyses identified key transcription factors and metabolic pathways governing seed storage. Additionally, the first comprehensive characterization of the *PEBP* gene family in faba bean identified 11 members, including allelic variation in *VfTFL1a*, a gene linked to determinacy. These findings provide novel genetic resources for marker-assisted breeding, advancing our understanding of nutrient accumulation and the genetic regulation of plant architecture. This work contributes to the development of high-performing faba bean varieties, supporting sustainable food and feed systems in northern climates.

Keywords: *Vicia faba*, GWAS, diversity panel, *TFL1*, seed development, determinacy, *PEBP*, RNA-seq.

Genetiska aspekter av fröutveckling, blomning och diversitet hos åkerböna

Abstract

Åkerböna (*Vicia faba* L.) är en näringsrik baljväxt med stor potential för hållbart jordbruk i nordliga klimat, men förädlingsframstegen har hittills varit begränsade. Denna avhandling kombinerar fenotypiska, genetiska och biokemiska analyser för att utveckla genomiska verktyg och påskynda förbättringen av viktiga agronomiska egenskaper och frökvalitet i åkerböna. En global diversitetspanel bestående av 220 accessioner, utvärderad under två odlingssäsonger, visade på en betydande fenotypisk variation i blomningstid, planthöjd, avkastning och frösammansättning. Inga negativa samband observerades mellan skörd och högt proteininnehåll, vilket tyder på möjligheten att förbättra båda egenskaperna samtidigt. En genome-wide association study (GWAS) identifierade 51 SNP-markörer kopplade till tio agronomiska egenskaper. Transkriptomiska och biokemiska analyser av fröutveckling avslöjade tydliga tidsmönster i ackumuleringen av näringsreserver. Proteiner ansamlades tidigt i både embryo och endosperm, medan stärkelsehalten ökade gradvis och oljehalten förblev låg. Genuttrycksanalyser identifierade centrala transkriptionsfaktorer och metaboliska vägar som reglerar fröinlagring. Vidare identifierades för första gången *PEBP*-genfamiljen i åkerböna, bestående av 11 medlemmar. Dessutom påvisades allelisk variation i *VfTFL1a*, en gen kopplad till determinans. Dessa resultat tillhandahåller nya genetiska resurser för markörbaserad växtförädling och ger en ökad förståelse av näringsackumulering i fröet, samt den genetiska regleringen av växtarkitektur. Arbetet främjar utvecklingen av nya åkerbönesorter och övergången till mer hållbara livsmedels- och fodersystem i nordligare klimat.

Genetische Einblicke in die Samenentwicklung, Blüte und Diversität der Ackerbohne

Abstract

Die Ackerbohne (*Vicia faba* L.) ist eine nährstoffreiche Leguminose mit großem Potenzial für nachhaltige Landwirtschaft in nördlichen Klimazonen. Allerdings sind die Züchtungsfortschritte bisher begrenzt. Diese Dissertation kombiniert phänotypische, genetische und biochemische Analysen, um genomische Werkzeuge zu entwickeln und die Verbesserung wichtiger agronomischer Eigenschaften sowie Samenqualität zu beschleunigen. Ein über zwei Saisons evaluiertes globales Diversitätspanel aus 220 Ackerbohnen-Akzessionen zeigt erhebliche phänotypische Variation in Blühzeit, Pflanzenhöhe, Ertrag und Samenzusammensetzung. Interessanterweise hatten höhere Proteingehalte keinen negativen Einfluss auf den Ertrag, was darauf hindeutet, dass beide Eigenschaften gleichzeitig verbessert werden können. Eine genomweite Assoziationsstudie (GWAS) identifizierte 51 SNP-Marker die mit zehn agronomischen Eigenschaften assoziiert sind. Transkriptomische und biochemische Analysen der Samenentwicklung offenbaren distinkte zeitliche Muster in der Akkumulation von Reservestoffen. Dabei akkumulieren Proteine früh im Embryo und Endosperm, während Stärke graduell zunimmt und der Ölgehalt niedrig bleibt. Genexpressionsanalysen identifizierten zentrale Transkriptionsfaktoren und metabolische Wege, welche die Nährstoffspeicherungen im Samen steuern. Zudem wurde erstmals die vollständige *PEBP*-Genfamilie in der Ackerbohne charakterisiert, bestehend aus 11 Mitgliedern. Allelische Variationen in *VfTFL1a*, einem Gen, das mit Determinanz assoziiert ist wurden nachgewiesen. Diese Ergebnisse stellen neue genetische Ressourcen für markergestützte Züchtung bereit und vertiefen unser Verständnis der Nährstoffakkumulation sowie der genetischen Regulation der Pflanzenarchitektur. Dies Arbeit trägt zur Entwicklung verbesserten Ackerbohnensorten bei und unterstützt den Übergang zu nachhaltigen Lebensmittel- und Futtersystemen in nördlicheren Klimazonen.

Preface

Most of the food we consume today has been profoundly shaped by breeding—a process influenced by human intervention and natural evolutionary forces, across millennia. Through domestication and subsequent selection, humans have gradually modified wild plants by favoring traits such as higher yield, disease resistance, and improved nutritional value, while selecting against characteristics like seed dormancy and shattering. This cumulative process of domestication, followed by increasingly systematic selection, has fundamentally transformed the crops that now form the backbone of our diets. Historically, early farmers and breeders relied on observation and empirical knowledge, selecting the most robust and productive plants from each generation to gradually enhance favourable traits. For example, modern corn (*Zea mays*) is the result of thousands of years of domestication and human selection from wild *teosinte* grass, with very little resemblance to the highly productive and nutritious crop we know today. Similarly, many fruits and vegetables—such as bananas and carrots—have undergone dramatic transformations from their wild progenitors, making them not only more delicious but also more suitable for large-scale cultivation. Natural evolutionary processes have also played a vital role in shaping our food crops. Random mutations and Darwinian selection, occurring alongside human intervention, have contributed to the remarkable diversity and resilience of agricultural species. The *Brassicaceae* family exemplifies this interplay, where a single wild species, *Brassica oleracea*, has given rise to an astonishing variety of cultivated crops. Through selective breeding, humans have developed cauliflower (selected for its arrested flower meristem), Brussels sprouts (developed from axillary buds), cabbage (enhanced terminal bud), and kale (modified leaf characteristics).

Addressing the current global challenge of feeding an ever-growing population requires a continuation and refinement of this breeding tradition. Agriculture, the foundation of global food production and a critical pillar of the global economy, is under increasing strain from climate change and the loss of arable land due to soil degradation and nutrient leakage. To meet these challenges, plant breeding has become more advanced, integrating high-throughput and data intensive technologies with traditional practices. Todays' plant breeding builds upon the foundational innovations of ancient agriculture, developed independently and in parallel in different regions and time points across the world. Among the earliest domesticated plants is faba bean (*Vicia faba L.*), a grain legume of significant cultural and nutritional value across diverse regions of the world. Renowned for its adaptability to varying climates and its ability to enhance soil fertility through nitrogen fixation, faba bean has long been a staple food source. Its potential to support sustainable agriculture lies in its role as a protein-rich crop that reduces the need for synthetic nitrogen fertilizers. However, modern breeding efforts for faba bean have lagged behind those for other major crops, limiting its broader adoption and agricultural impact.

This thesis aims to address this gap by focusing on pre-breeding strategies for faba bean, with particular emphasis on characterizing its phenotypic diversity and examining key developmental traits. Specifically, it seeks to answer critical questions such as: When during seed development are nutrients stored? Can variations in plant architecture be attributed to specific genes? How do hundreds of different faba bean accessions from across the globe perform in the fertile soils of southern Sweden? Can we retrieve informative genetic markers that show an association with agriculturally relevant traits? By exploring these questions and presenting new insights, this thesis aspires to contribute meaningfully to the advancement of a historically significant crop, offering a foundation for its future improvement in the face of global agricultural challenges.

Dedication

To my family, who fill this journey with meaning:
To my daughters, Ingrid Vilde and Wilma Lo, born alongside these pages.
To the two generations of female scientists before me, Elinor and Heidi, who paved the bumpy road for me to dance on; and to Marlis, who instilled in me a love for plants.
To my father, Claus, my radical reminder of the curiosity at the heart of it all.
To Ellen and Jens, my anchors.

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Ohm, H**., Saripella, G.V., Hofvander, P. & Grimberg, Å. (2024). Spatio-temporal transcriptome and storage compound profiles of developing faba bean (*Vicia faba*) seed tissues. Frontiers in Plant Science, 15. https://doi.org/10.3389/fpls.2024.1284997.

II. **Ohm, H**., Åstrand, J., Ceplitis, A., Bengtsson, D., Hammenhag, C., Chawade, A., & Grimberg, Å. (2024). Novel SNP markers for flowering and seed quality traits in faba bean (*Vicia faba* L.): characterization and GWAS of a diversity panel. Frontiers in Plant Science, 15. https://doi.org/10.3389/FPLS.2024.1348014

III. **Ohm, H**., Mahmood, U., Östberg, J., Alverup, J., Grimberg, Å., Hofvander, P. (2025). Comprehensive Identification of the *PEBP* Gene Family in Faba Bean (*Vicia faba* L.): Insights into *VfTFL1a* Variability and Growth Determinacy. Submitted.

The contribution of Hannah Ohm to the papers included in this thesis was as follows:

I. Conducted all laboratory and greenhouse work, including the methodological development, and preparation of samples for RNA sequencing. Analysed the curated bioinformatics data and contributed to the formal analysis. Drafted the original manuscript and participated actively in the review and editing process.

II. Participated in all aspects of planning and performing the field trials, data collection, phenotypic data curation, investigation and methodology development. Drafted the original manuscript with co-authors as well as contributed to the review and editing process.

III. Conducted laboratory and field work, developed methodology, performed formal sequence analysis and investigation. Curated parts of the retrieved bioinformatics results, and wrote the original draft of the manuscript.

List of figures

Abbreviations

ANF	Antinutritional factor
FFA	Free fatty acids
DET	Differentially expressed transcript
GBS	Genotyping-by-sequencing
GWAS	Genome-wide association studies
GxE	Gene-environment interaction
LAFL	LEAFY COTYLEDON 1, ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON 2
LD	Linkage disequilibrium
MAF	Minor allele frequency
PEBP	Phosphatidylethanolamine-binding proteins
QTL	Quantitative trait loci
RNA-seq	RNA sequencing
SDG	Sustainable Development Goals
SNP	Single nucleotide polymorphism
TAG	Triacylglycerol
TF	Transcription factor
TFL1	Terminal Flower1
WRI1	WRINKLED1

1. Background

1.1 Legumes for sustainable agriculture

As the global population grows and environmental challenges become more urgent, cultivating sustainable, nutritious, and locally produced sources of food is increasingly critical. In this context, legumes stand out as versatile crops with the potential to contribute significantly to several Sustainable Development Goals (SDGs) (Figure 1). These include eradicating hunger (SDG 2), promoting health and well-being (SDG 3), and, in alignment with climate action, protecting terrestrial ecosystems (SDG 12-15).

Figure 1 UN Sustainable Development Goals. Non-shaded areas highlight topics of specific relevance for the increased cultivation of legumes (UN 2025).

Transitioning to plant-based proteins in our diets is crucial for minimising the environmental footprint of protein production and consumption, supporting the nutritional needs of a growing global population, and

addressing the negative consequences of industrial livestock farming (Duluins & Baret 2024). This dietary shift could offer numerous advantages linked to food security, nutrition, and sustainable agriculture - aligning with SDG 2, which aims to decrease the estimated 2 billion people in the world without regular access to safe, nutritious, and sufficient food (UN 2020; FAO 2024). Malnutrition can have various faces; while 820 million people suffer from direct hunger, 2 billion adults and 40 million children (<5 years) are overweight (UN 2019). Pulses (grain legumes), the dry edible seeds of legumes, are critical to addressing these dual challenges. As the second-most important crop group after cereals, legumes provide a high-quality protein source that complements the essential amino acids found in cereals, making them vital for combating malnutrition (Sharma *et al.* 2013). In addition, many legumes contain high amounts of dietary fibre, resistant starch, and bioactive compounds that contribute to improved health outcomes, aligning with SDG 3 (Willett *et al.* 2019; Ferreira *et al.* 2021).

Legumes also play a vital role in enhancing biodiversity and soil health. Agricultural biodiversity is defined as the genetic, species, and ecosystem variation within food systems - both cultivated and wild (FAO 2019). The intensification of agriculture has led to less diversity in crop rotations and higher genetic uniformity among crop species (Hufnagel *et al.* 2020). Currently, of the approximately 6,000 plant species cultivated for food worldwide, fewer than 200 play a significant role in the global food supply. Among these, just nine account for 66% of total crop production, while only three—wheat, rice, and maize—collectively provide the bulk of global dietary energy intake (FAO 2019, 2025). This dominance of a few highly produced crops in our current industrial agriculture dates back to the promotion of food self-sufficiency after the Second World War (Mawois *et al.* 2019). Today a limited number of actors influence a majority of the food system governance (IPES-Food 2016). High-input systems and strict market standards, coupled with consumer demand for inexpensive, uniform and predictable food have further reinforced the few, large breeding and production companies on multinational levels with unproportioned influence on the agricultural system. Food, like any other commodity, is increasingly traded, with an expansion of roughly 350% in the global food trade from 2000-2021 (UNCTAD 2024). By contrast, SDG 2 calls for sustainable food production, resilient agriculture, and well-functioning food markets to reduce the vulnerability and price fluctuations of staple crops, which can be

impacted by import dependence. Consequently, countries must carefully balance national food sovereignty with imported food to minimize climate risks while ensuring long-term food security (Mbow *et al.* 2019). The European Union provides a lucrative market for non-European farmers, as it is a major consumer and importer of pulses. Over the past five years, the European pulses market has grown by more than 27%, and the area dedicated to cultivating protein crops within Europe is projected to increase by 37% during 2020-2030 (EC 2020).

Beyond the narrowing of crop diversity, industrialized agriculture has contributed to uniform landscapes, eutrophication, nitrate pollution of waterways, and soil degradation. To reverse this development, and in line with SDG 15, three strategies stand out: diversifying crop rotations or using intercropping, reintroducing minor crops into production systems, and enhancing genetic diversity within key crop species. Legumes align well with these approaches, offering agronomic benefits such as improved soil fertility, increased subsequent crop yields, disease-cycle disruption, and reduced reliance on synthetic fertilizers (Jensen *et al.* 2010; Ballard 2020; Zhao *et al.* 2022). Moreover, many legumes also attract pollinators and serve as refuges for other wildlife (EC 2018).

Today, food systems are responsible for 21-37% of the global anthropogenic greenhouse gas emissions, with nitrous oxide (N_2O) emissions from fertiliser applications and livestock contributing majorly (Mbow *et al.* 2019; Crippa *et al.* 2021). Synthetic fertilizers—primarily nitrogen (N), phosphorus (P), and potassium/potash (K)—greatly amplified agricultural yields over the past half-century. However, crops assimilate only about half of the 109 million metric tons of fertilizer nitrogen applied annually, leaving the remainder to leach or volatilize as greenhouse gases (Lassaletta *et al.* 2014; Peoples *et al.* 2019). Agronomically, legumes deliver critical benefits to cropping systems by being self-sufficient in their nitrogen supply due to their ability to fix atmospheric nitrogen in symbiosis with the soil-borne bacteria *Rhizobium.* Not only does this eliminate the need for synthetic nitrogen fertilisation throughout the entire cultivation period, but it also enhances soil microbial activity and nutrient availability for the subsequent crop, particularly when incorporated as green manure or when stem, leaf, and root residues are retained in the field after harvest (De Pascale *et al.* 2018). Without affecting the yield levels, N fertilisation rates for cereals grown after grain legumes can be reduced by roughly one-third (Röös *et al.*

2020; FAO 2025). Consequently, including legumes in the crop system reduces the use of fertilisers which in turn limits greenhouse gas emissions (FAO 2019), thereby meeting climate objectives under SDG 12-15.

1.2 Faba bean: A legume for northern climates

1.2.1 Agronomic performance and seed quality traits

One legume species with exceptional potential for advancing these sustainability goals in a northern context is *Vicia faba,* commonly referred to as broad bean, horse bean, Celtic bean, field bean, fava bean, or faba bean (Figure 2). Faba bean has the highest nitrogen-fixing capacity among cool-season legumes (Jensen *et al.* 2010) and represents a valuable "minor crop" with untapped potential for diversifying agricultural systems. Its high genetic diversity provides opportunities for breeding resilient varieties adapted to varied environments and cultivation practices. The adaptability of faba bean to cooler climates makes it particularly valuable in regions with lower temperatures, where other legumes, such as soybean, are less suited for cultivation. By that, it offers a local alternative to soybean, whose cultivation is resource-intensive and often geographically distant from markets. As already indicated by its historical denomination "the poor man's meat", the "green proteins" from faba bean can serve well as a supplement to, or replacement of, the "red proteins" from meat (Cooper *et al.* 2017; Warsame *et al.* 2020). Moreover, national policy initiatives such as Sweden's National Food Strategy (2016) emphasize enhanced self-sufficiency for domestically produced foods (Regeringen 2017). Even though legumes such as faba bean and peas, the latter currently cultivated to a relatively large extent in the region, have been traditionally grown in Sweden and the rest of Europe (Leino 2023), these countries are today highly dependent on large imports of especially soybean for mainly animal feed (EC 2018, 2024).

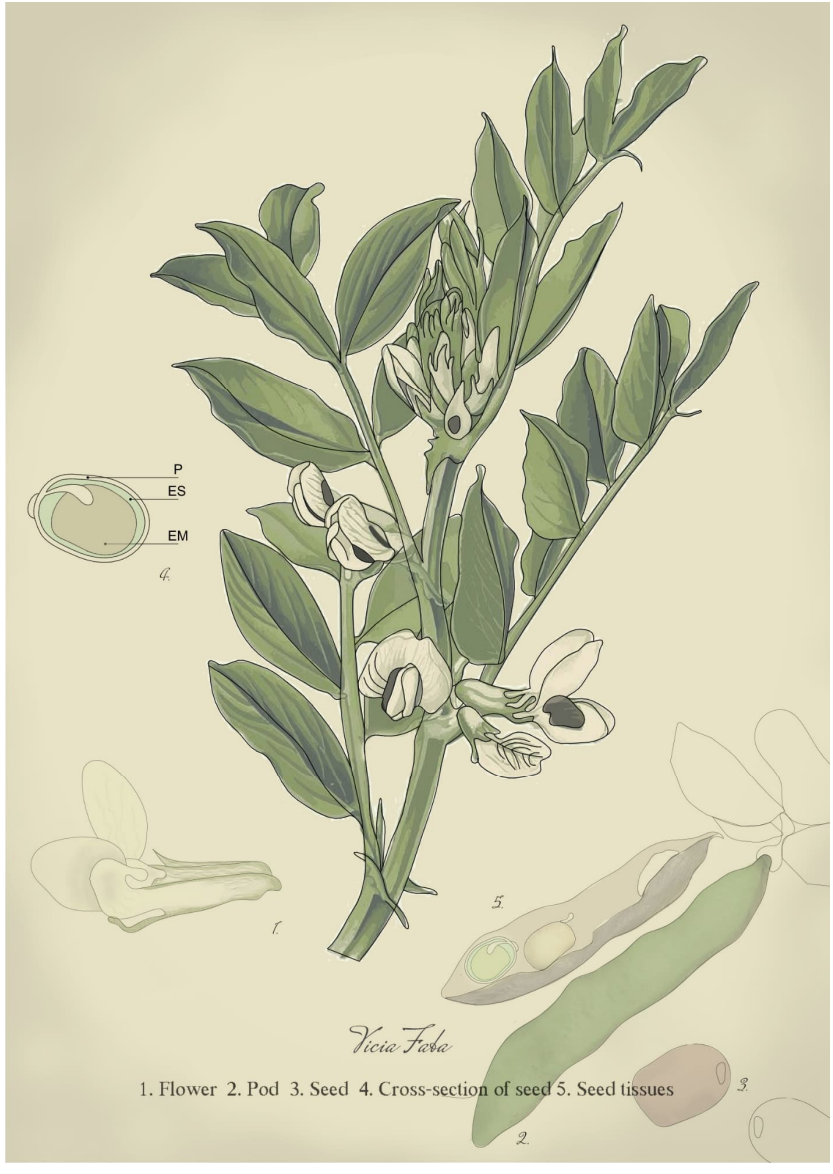


Figure 2 Botanical illustration of the faba bean plant, an annual legume crop belonging to the *Fabaceae* family, characterized by its erect growth habit, often with variegated flowers (white flowers with a distinctive black spot on the wing petals) and pinnately compound leaves, with 2-7 leaflets. Included in the picture is a cross-section of the faba bean seed, depicting its different tissues: pericarp (P), endosperm (ES), and embryo (EM). Illustration by E. Rémy, created for this thesis.

Compared to other legumes (Table 1), the faba bean stands out for its high seed protein, low oil content, and low allergenicity (Smits *et al.* 2021). The seeds are rich in carbohydrates with a low glycemic index and high levels of resistant starch and dietary fibre, offering health benefits and making them suitable for human consumption and animal feed (Ellwood *et al.* 2008; Multari *et al.* 2015; Chetto *et al.* 2024). Faba bean contain bioactive compounds as antioxidants, phenolic compounds, and γ-aminobutyric acid, supporting cardiovascular health and reducing the risk of diabetes and cancers (UN 2017; Valente *et al.* 2019). Additionally, L-DOPA in faba bean has therapeutic applications for Parkinson's disease treatment (Multari *et al.* 2015).

Table 1 Comparison of legumes' nutritional content. Values are given per 100 g of raw product. Data sources: NFSA (2024); values marked with [1] are sourced from USDA (2019).

Legume species	Energy (kcal)	Protein (g)	Fat (g)	Carbohydrates (g)	Dietary fibre (g)
Faba bean	316	25	1.7	42	17
Soybean	384	36	19	10	16
Yellow pea	334	22	2.5	50	13
Brown bean	313	22	1.5	45	16
Chickpeas[1]	378	20.5	6	63	11
Green lentils	263	20.5	2	29	23

Despite these advantages, faba bean also contains antinutritional factors (ANFs) such as tannins, which can negatively impact protein digestibility, and vicine-convicine, which pose severe health risks for individuals with the genetic disease called favism. Advances in plant breeding have led to low-vicine-convicine varieties and low-tannin varieties, improving their suitability for consumption (Crépon *et al.* 2010; Khamassi *et al.* 2019).

1.2.2 Historical context

Faba bean has been a crucial crop for human and animal nutrition for millennia, with possible domesticated varieties dating back approximately 10.000 years ago (Tanno & Willcox 2006). Although the crop's wild

ancestor remains unidentified, evidence indicates that wild forms of *Faba bean* were gathered and consumed during the Natufian period (12,500–9,500 BCE) in what is now Israel (Caracuta *et al.* 2016). From there, the species gradually adapted across the Mediterranean basin (Valamoti 2023). By antiquity, faba bean had become a staple food but was also subject to cultural beliefs and ritual practices, exemplified by Pythagoras' reputed ban on its consumption (Figure 3) and the Roman Lemuria rituals. One reason behind these prohibitions may have been the prevalence of favism—a genetic disorder in humans, relatively common in the Mediterranean region, that causes severe hemolytic reactions to faba bean consumption. Yet the ritual significance of faba bean also reflects a mix of medical, cultural, and botanical factors extending beyond the disease itself (Matias 2024). Today, over 38.000 accessions of faba bean are maintained in global genebanks (Duc *et al.* 2010), preserving its genetic diversity for pre-breeding and breeding programs.

Figure 3 Historical depiction of Pythagoras' ban on faba bean, "Do Not Eat Bean" [Fol. 25 Recto], 1512/1514. Image courtesy of the National Gallery of Art, Washington D.C.

1.2.3 Global production of faba bean

Globally, faba bean is the sixth most widely grown grain legume (FAO 2025). Despite substantial population growth and rising food demand since the 1960s, the global total cultivated area of faba bean has declined - largely attributed to the widespread adoption of synthetic fertilizers, narrow profit margins, and commercial dominance of soybean (Mínguez & Rubiales 2021). As a result, faba bean's current annual production remains relatively modest at approximately 6 million metric tons, covering 2.7 million hectares (Figure 4A), and the crop's significance is minor compared to soybean, which is the leading grain legume grown on 120 million hectares (Ballard 2020; FAO 2025). Major production areas of faba bean (Figure 4B) include Asia, Europe, and Africa with key producers China, Ethiopia, Australia, and the UK (Cooper *et al.* 2017; Aguilar-Benitez *et al.* 2021; FAO 2024). Faba bean is utilised in various food applications in many different cultures, including as an ingredient in bread, pasta, and meat alternatives. While primarily used as animal feed in Europe and the Americas, faba bean is a staple protein source in China and North Africa (Ballard 2020).

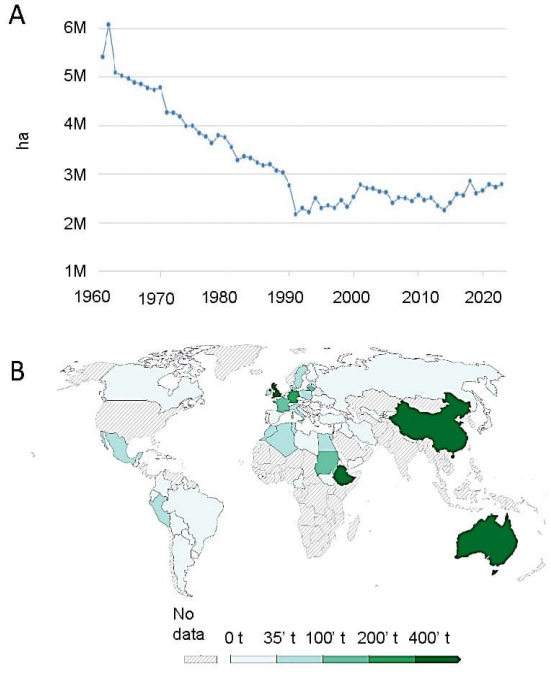


Figure 4 Global faba bean (dry) cultivation and production. A. World harvested area (ha) during 1960-2020 (FAO 2025). B. World view of production (tons) in 2022 (FAO 2024).

Faba bean is a cool-season annual legume, sown in both spring and autumn in warm-temperate regions, but only in spring in our northern climate (Jensen *et al.* 2010). Winter types, sown in the fall, utilise autumn and winter moisture, mature earlier and yield more than the spring types due to their efficient utilisation of the growing season. Although frost tolerant, autumn-sown varieties of faba bean face challenges such as frost damage and shorter growing periods, particularly in northern European regions (Link *et al.* 2010; Stoddard & Hämäläinen 2011). However, they are frequently cultivated in the UK and southern parts of Europe, where milder winter conditions support their growth and productivity. Breeding programs for winter types suitable for harsher climates have seen limited progress since the 1980s, hindered by the risk of underdeveloped plants during freezing periods (Link *et al.* 2010; Inci & Toker 2011; Flores *et al.* 2012; Landry & Hu 2019).

1.2.4 Recent advances and key breeding objectives

The genomic research for faba bean have lagged behind those of most other major grain legumes (Adhikari *et al.* 2021). Faba bean possesses a massive and complex genome of approximately 13 Gb, distributed across six chromosome pairs (2n = 12). This substantial size is attributed to a high density of repetitive elements, transposable sequences, and extensive intergenic regions (~330 kb) (Jayakodi *et al.* 2023). By comparison, other legumes such as *Pisum sativum* (pea) and *Glycine max* (soybean) have notably smaller genomes, at approximately 4.3 Gb and 1.1 Gb, respectively. Early genotyping arrays for faba bean were often derived from limited transcriptome datasets, potentially introducing ascertainment bias (Khazaei *et al.* 2021). Recent milestones, however, include the publication of a high-quality, chromosome-level reference genome (Jayakodi *et al.* 2023), which is vital for dissecting gene function, evolutionary relationships, and trait inheritance (Bernal-Gallardo & de Folter 2024). A faba bean pan-genome study is also underway, aiming to capture more extensive intraspecies diversity (Auvinen *et al.* 2023).

Adding to the genetic complexity is the faba bean's partially allogamous reproductive system, combining self-pollination (autogamy) and cross-pollination (allogamy) (Link 1990). The resulting high heterozygosity, meaning that an individual organism carries different versions (alleles) of a gene at a specific location on its DNA, presents both challenges and opportunities for breeders. On the one hand, it complicates the development

and maintenance of pure lines, which are crucial for uniformity in breeding programs. On the other hand, it provides a broader genetic base for selection and opportunities to exploit partial or full heterosis, a phenomenon in which the hybrid offspring surpass their parental lines, in yield or resilience (Bishop & Nakagawa 2021). In faba bean however, heterosis is notably limited to partial heterosis due to the absence of cytoplasmic male sterility systems, which prevents the development of true hybrid cultivars (Brünjes & Link 2021). Production challenges such as susceptibility to frost, high temperatures, drought stress during flowering, and pests or diseases further impede faba bean's widespread cultivation (Aguilar-Benitez *et al.* 2021; Abou-Khater *et al.* 2022). Common issues in Europe include chocolate spot disease, a foliar disease caused by *Botrytis fabae* and *Botrytis cinerea*, which can lead to severe yield losses, and bean weevil (*Bruchus rufimanus*) infestations, which damage seeds and contaminate the harvest with beetle remnants (Stoddard *et al.* 2010; Huber *et al.* 2023).

A list of breeding objectives and recent advancements in faba bean research is presented in Table 2, highlighting efforts to enhance key traits such as higher and more stable yields, optimising seed protein composition to meet dietary needs, reducing ANFs, enhancing tolerance to abiotic stresses like drought and frost for climate resilience, and improved plant architecture (e.g., stronger stems, terminal growth, and optimal pod height). Progress in these areas will enhance the competitiveness of faba bean within the rapidly expanding plant-protein sector. In Sweden, faba bean breeding programs were active between 1930 and 1990 (Svalöf) and introduced cultivars tailored to local conditions, focusing on early maturing and improved seed shapes and sizes for mechanised sowing and harvesting, such as *Primus*, *Sving*, *Aurora*, and *Arla* (Lyhagen 2016). Between 1956 and 1967, the Swedish breeder Jan Sjödin developed a comprehensive faba bean mutation population, yielding numerous mutants with distinct seed, flower, and plant-architecture traits (Sjödin 1971; Lyhagen 2015; Grimberg 2019; Khazaei *et al.* 2024). Although Swedish-based breeding ceased in the 1990s, crucial genetic material remains conserved at NordGen (the Nordic Genetic Resource Center). In 2019, Lantmännen, a Swedish agricultural cooperative, initiated a new breeding program targeting faba bean, signalling a renewed interest in domestic varietal development.

Table 2 Breeding objectives and current advances in faba bean research

Trait Amelioration	Current Advances	Selected references
Higher/stable yields	Identification of yield-related markers and varieties with consistent yields across environments.	Cruz-Izquierdo *et al.* 2012; Ávila *et al.* 2017; Skovbjerg *et al.* 2020; Adhikari *et al.* 2021; Gela *et al.* 2023; Gutierrez *et al.* 2024.
Reducing ANFs	Development of varieties with reduced levels of anti-nutritional factors through elucidation of their biosynthetic pathways, enhancing digestibility, nutritional value, and palatability.	Crépon *et al.* 2010; Webb *et al.* 2016; Gutierrez & Torres 2019; Zanotto *et al.* 2020; Björnsdotter *et al.* 2021; Lippolis *et al.* 2025.
Seed nutrients	Enhancement of protein content, amino acid profiles, and identification of quality-related markers to improve seed nutritional composition.	Warsame *et al.* 2020; Martineau-Côté *et al.* 2022; Zhao *et al.* 2023.
Improved plant architecture	Selection for agronomically favourable traits including branching, lodging and determinate growth patterns.	Ávila *et al.* 2017; Nurmansyah *et al.* 2019; Hughes *et al.* 2020a.
Biotic and abiotic stress tolerance	Development of varieties with enhanced tolerance to drought, salinity, heat, and frost to ensure adaptation to climate variability. Identification of resistance sources against major foliar diseases, root rot pathogens, and parasitic weeds (*Orobanche* spp.).	Abbes *et al.* 2007; Link *et al.* 2010; Lavania *et al.* 2015; Khan *et al.* 2019; Sallam & Ul-Allah 2019; Yang *et al.* 2020; Faridi *et al.* 2021; Ijaz *et al.* 2021; Maalouf *et al.* 2022; Segers *et al.* 2022; Gutiérrez *et al.* 2023; Webb *et al.* 2024.
Earliness	Development of early-maturing varieties.	Stoddard & Hämäläinen 2011; Aguilar-Benitez *et al.* 2021.
Herbicide resistance	Development of resistant varieties to multiple herbicides with different modes of action.	Mao *et al.* 2019; Abou-Khater *et al.* 2022.

1.3 The importance of pre-breeding for crop improvement

Pre-breeding serves as a crucial approach for addressing key challenges in faba bean breeding by introducing valuable genetic variation from diverse sources, including older varieties and landraces. These genetic resources may contain traits essential for addressing breeding limitations. In essence, pre-breeding serves as a critical link in modern agriculture, bridging valuable genetic resources with elite breeding lines to enhance crop performance. Despite its significance, pre-breeding remains scarcely described in the literature as a distinct scientific field or theoretical framework. Nevertheless, it constitutes a fundamental component of agricultural science, underpinning both basic research and practical applications. Moreover, it plays a pivotal role at the science-policy interface by facilitating the mobilisation of scientific expertise for sustainability. The need for such scientific engagement has become increasingly urgent, particularly in the context of the United Nations' 2030 Agenda for Sustainable Development (UN 2019).

Pre-breeding involves the identification and incorporation of desirable traits from diverse germplasm, which can then be introduced into breeding pipelines (Figure 5). This process enhances genetic diversity while ensuring the development of robust populations with traits that meet current or future agricultural challenges (Sharma *et al.* 2013; Kailash *et al.* 2017; Sukumaran *et al.* 2022). The characterization of germplasm that can act as donor populations for these traits provides a critical foundation for addressing long-term breeding objectives. These objectives often involve trade-offs or competing priorities, such as balancing resilience to climate change, pest resistance, improved nutritional profiles, and local climate adaptations (Allier *et al.* 2020).

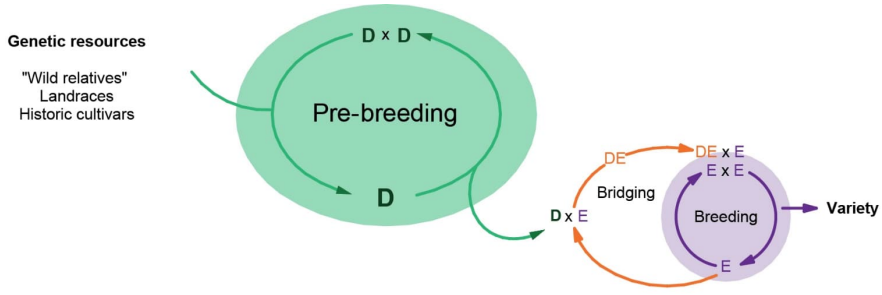


Figure 5 Schematic representation of pre-breeding in relation to breeding. The diagram illustrates the flow from diverse donors (D) to elite material (E) in the breeding process. Modified from Allier et al. 2020, with the authors' permission.

Modern breeding programs integrate classical genetics with statistical computation and molecular biology, identifying promising accessions by progressively narrowing selection over many generations. Pre-breeding strengthens these efforts by identifying, evaluating, and transferring beneficial alleles from diverse germplasm into pre-elite lines, which can then be used in breeding programs. By increasing genetic variation, pre-breeding enhances directly the genetic gain obtained through selection, which is critical for crop improvement. The classical Breeder's Equation, here in its refined version, $R = ir\sigma_A$, serves as a fundamental model estimating the response to selection (R) which predicts the expected improvement in a trait across generations, reflecting how effectively genetic parameters drive change within a population (Falconer & Mackay 1996). Here, R represents the response to selection, i is the selection intensity, which measures the standardized selection differential, r denotes selection accuracy, which measures the correlation between estimated and true breeding values, and σ_A is the additive genetic standard deviation, derived as the square root of the additive genetic variance (V_A). Pre-breeding primarily contributes to this equation by expanding the additive genetic variance and consequently the additive genetic standard deviation (σ_A), thus enhancing the potential for genetic gain in breeding programs. However, a greater genetic diversity could also make it harder to get precise heritability estimates, due to more gene-environment interactions (GxE), and more epistasis (gene-to-gene interactions). Additionally, careful management is required to balance short-term genetic losses (resulting from maladaptive traits) with long-term benefits such as increased adaptability and resilience (Bassi *et al.* 2024).

While breeding programs usually prioritise high selection intensity for short-term goals, pre-breeding programs adopt a longer-term perspective, focusing on maintaining and leveraging genetic diversity to support future breeding efforts and act as the foundation for elite breeding populations (Allier *et al.* 2020; Mackay 2020). Breeding inherently reduces genetic diversity due to its focus on creating uniform lines and applying narrow selection criteria. Yet, it is the genetic diversity that is the raw material for plant breeders to work with. To secure and understand the available genetic diversity, pre-breeding is therefore essential. This is particularly relevant for crops such as faba bean, where the available germplasm is restricted to cultivated forms, exhibiting extensive genetic diversity across subspecies, landraces, and varieties but lacking a wild-type progenitor to introduce natural variation (Duc *et al.* 2010). Consequently, all existing genetic resources, including older and less agronomically competitive varieties, become invaluable reservoirs of traits that may prove critical for future breeding efforts (Hayward 1993). These resources allow breeders to identify rare alleles which may otherwise be lost in the process of narrowing down to a few market-ready varieties (IPES-Food 2016). Furthermore, they support efforts toward regionally adapted crop varieties, facilitating more specific breeding targets aligned with local growing conditions.

For faba bean, prioritising pre-breeding not only enhances the crop's role in sustainable food systems but also unlocks its potential to address the nutritional and environmental needs of a growing global population. With only a slight increase in domestic cultivation—from 2.2% to 3.2% of arable land—Sweden could replace up to 50% of its meat consumption with domestically grown grain legumes, according to Röös *et al.* (2020). This modest cultivation increase would deliver significant benefits, including health benefits and a 20% reduction in the climate impact of our food (Röös *et al.* 2020). However, achieving these goals will require the development of new crop varieties adapted to cultivation in Sweden, with the desired yield stability and seed quality.

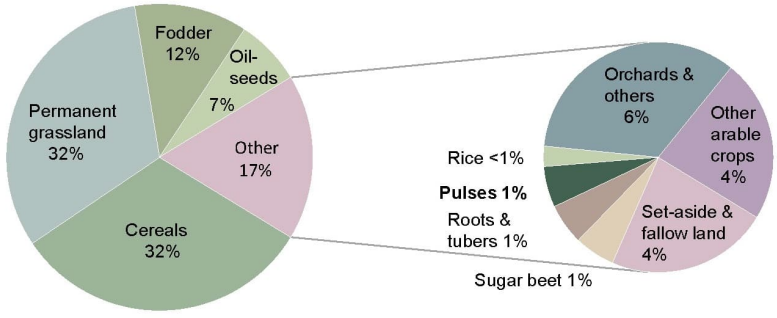


Figure 6 Land use in EU 2023, figure adapted from EC (2024).

Despite the clear benefits, the cultivation of grain legumes remains historically low in Europe, occupying merely 1% of the total agricultural land (Figure 6), with the majority allocated for animal feed (EC 2024). This underutilisation underscores the need for pre-breeding and breeding efforts to facilitate the integration of legumes into agricultural systems. A global dietary shift from animal-dominated to plant-dominated protein sources is widely recognised as essential for the environmental sustainability of our food system. However, the expansion of legumes in the food sector is hindered by an underdeveloped value chain, encompassing gaps in breeding initiatives, harvest management, processing infrastructure, and market development. As a result, the full potential of legumes as sources of "green protein", "whole foods", and "locally produced crops" remains largely unrealised.

2. Aims and objectives

This thesis aims to facilitate the improvement of faba bean for Nordic environments and sustainable agriculture by characterising available phenotypic diversity, identifying critical genetic determinants of agronomic important traits, and providing new knowledge that can enhance future breeding efforts.

Objectives:

I. Characterise the genetic and phenotypic diversity of a faba bean diversity panel in a Nordic climate, assessing essential agronomic and seed quality traits.

II. Enable the development of genomic breeding tools: Identify genetic markers and molecular pathways associated with key agronomic traits through GWAS and transcriptomics.

III. Increase the understanding of genetic regulation of key traits: Investigate the molecular regulation of flowering and plant architecture, with a specific focus on the *PEBP* gene family and determinacy.

By achieving these objectives (Figure 7), this thesis aims to contribute to the development of environmentally adaptive and high-performing faba bean varieties, addressing the global demand for sustainable and nutritious plant-based protein sources.

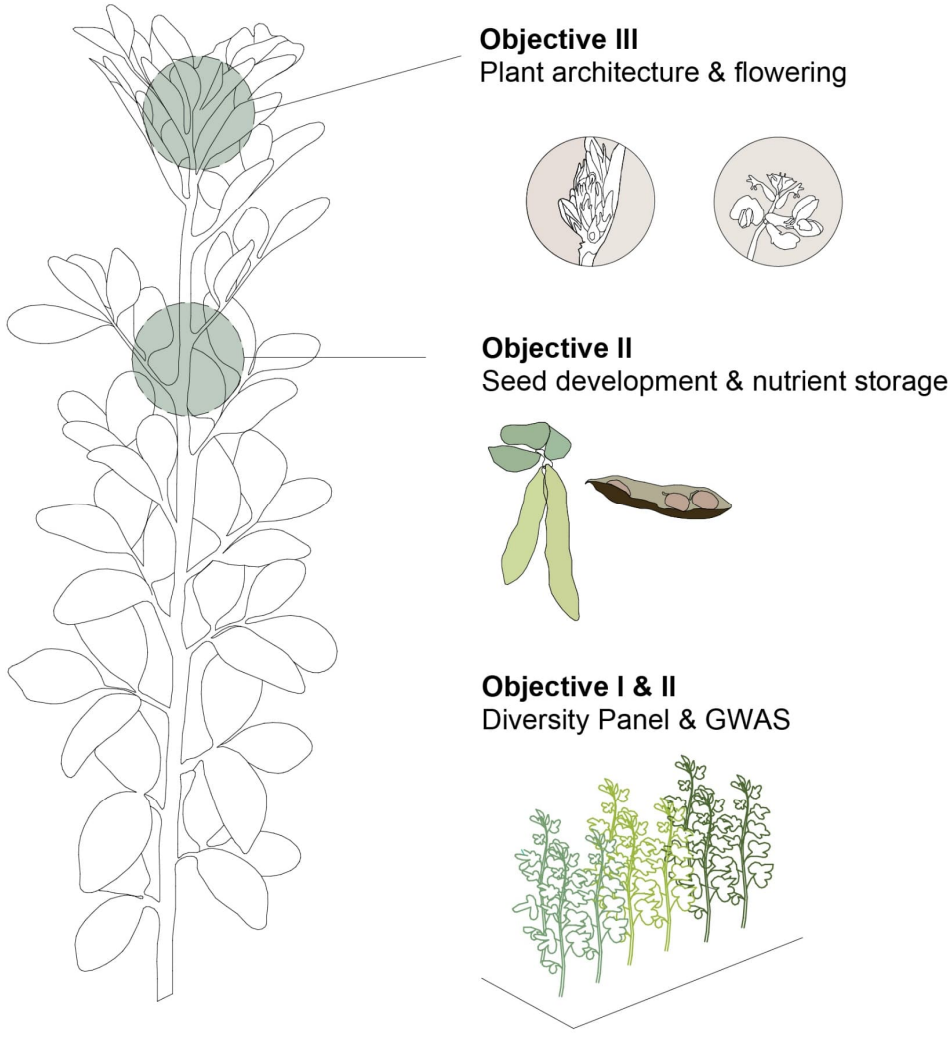


Figure 7 Schematic representation of the main research themes, the thesis objectives and their connection to the pre-breeding of faba bean. Illustration by E. Rémy.

3. Faba bean germplasm diversity

To effectively characterise the genetic resources of faba bean and assess their agronomic performance in Nordic climates (**Objective I**), a diverse collection of accessions must be selected. This diversity panel serves as the foundation for evaluating genetic and phenotypic variation and identifying accessions that carry traits relevant to environmental adaptation and agronomic performance. Furthermore, it provides the basis for identifying genetic markers and molecular pathways associated with these traits through genome-wide association studies (GWAS) and transcriptomics, thereby supporting the development of genomic breeding tools (**Objective II**). The following sections outline the methodology and structure of this study, beginning with a description of the diversity panel composition and field trial design. This is followed by key findings from **Paper II**, including the extent of phenotypic variation observed and the results from GWAS based on the field trials and genotyping.

3.1 Diversity panel composition and field trial design

Representing a wide range of variation within a plant species, a diversity panel is a collection of genetically diverse accessions typically including landraces, wild relatives, commercial cultivars and elite breeding lines. The faba bean diversity panel established in this project was sourced from gene bank material and commercially available cultivars, covering a broad geographical distribution across all continents, with the majority originating from northern and central Europe. Accessions were selected to represent different breeding advancement statuses and were evenly distributed among the four groups, arranged from least to most advanced; cultivated (including those of unknown origin), landraces (including known heirloom cultivars),

advanced breeding or research lines, and registered varieties. Additional inclusion criteria for the diversity panel were accessions with low ANFs, a range of plant heights, determinate growth types, diverse flower colours, and a variation of seed sizes, based on information from their providers.

Generally, the number of accessions in a diversity panel to be used in a GWAS should be chosen to ensure sufficient statistical power for identifying genetic markers linked to target traits and heritable phenotypes (Sukumaran *et al.* 2022). Approximately 250 accessions were initially selected and grown in both field and greenhouse conditions during 2020. This amplification phase allowed for evaluation of basic agronomic traits and seed production capacity, leading to the final selection of 220 accessions suitable for the field trials. These accessions formed the basis for field trial characterisation and GWAS (Paper II), with a subset selected for more detailed molecular studies (Paper III) described in the next chapter. To address challenges associated with partial allogamy during seed amplification in field, up to 15 plants per accession were isolated in bags to prevent cross-pollination between different accessions (Figure 8A). Field trials of the diversity panel were conducted in southern Sweden (55.65_ N, 13.06_ E) in 2021 and 2022, using an alpha-lattice structure in two replicates (i.e. 2 x 220 plots each year) (Figure 8B). Given the large number of accessions, an alpha-lattice design was chosen to minimise intra-block variation and improve statistical power, as recommended for field trials (Akinwale *et al.* 2021). Each plot, approximately 1–2 m² in area, was sown at a density of 50 seeds per square meter. Due to limited seed availability, seeds harvested from the 2021 field trial were used for sowing in 2022, while seeds from the initial round of amplification (i.e. from isolated plants) were used for genotyping.

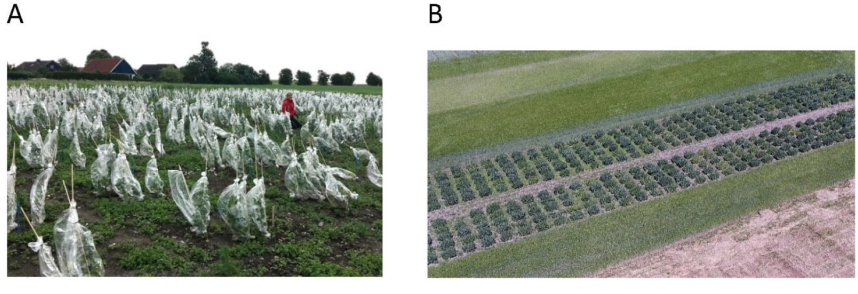


Figure 8 Field trial design and pollination control. A. Plants isolated with bags to prevent cross-pollination between accessions during the seed amplification in the field at Svalöv, 2020. Field trial setup in Lönnstorp, 2021. Photos: A. Chawade and Å Grimberg. B.

3.2 Agronomic trait variation in a Nordic climate

To assess the diversity of agronomic and seed quality traits of faba bean under Nordic climate conditions, a wide range of traits were phenotyped among the 220 accessions in our diversity panel in field trials (Table 3).

Table 3 Phenotypic measurements in field trials and post-harvest

Trait	Method
Establishment	% Established plants measured at ~ 35 days after sowing.
Flowering	Days to flowering: Time from sowing until >50% of plants per plot showed ≥1 open flower.
Height	Average height of 5 plants/plot, measured at ~76 days after sowing.
Maturity	Days to maturity. Time from sowing until >50% of plants have filled, black/dry pods.
Thousand Grain Weight	Weight of 1000 seeds (g)
Yield	Weight of seeds/plant (g)
Size	Seed area
Bean Weevil Damage	The proportion of infested seeds assessed post-harvest.
No Seeds/Plant	Number of seeds per plant
Protein Content	% Protein of seed dry weight
Starch Content	% Starch of seed dry weight

Traditional methods of phenotyping of plants are time-consuming, labour-intensive, and prone to human errors. While manual phenotyping (visual assessment and scoring) was primarily used in our field study of the diversity panel, drone-based imaging and on-site cloud data handling were also tested. Comparisons of drone-derived metrics—including establishment, height, plants per square meter, ground coverage, vegetation indices VARI (Visible Atmospherically Resistant Index) and ExG (Excess Green Index), senescence, and chlorophyll content—with manual measurements or drone data for related traits showed strong correlations for several parameters. In

particular, drone and manual measurements of plant height late in the season, as well as drone-based assessments of chlorophyll content and senescence, demonstrated high agreement (Figure 9). These findings highlight the potential of drone-based methods for faba bean screening in field to reduce labour demands while ensuring high-quality data collection for these traits.

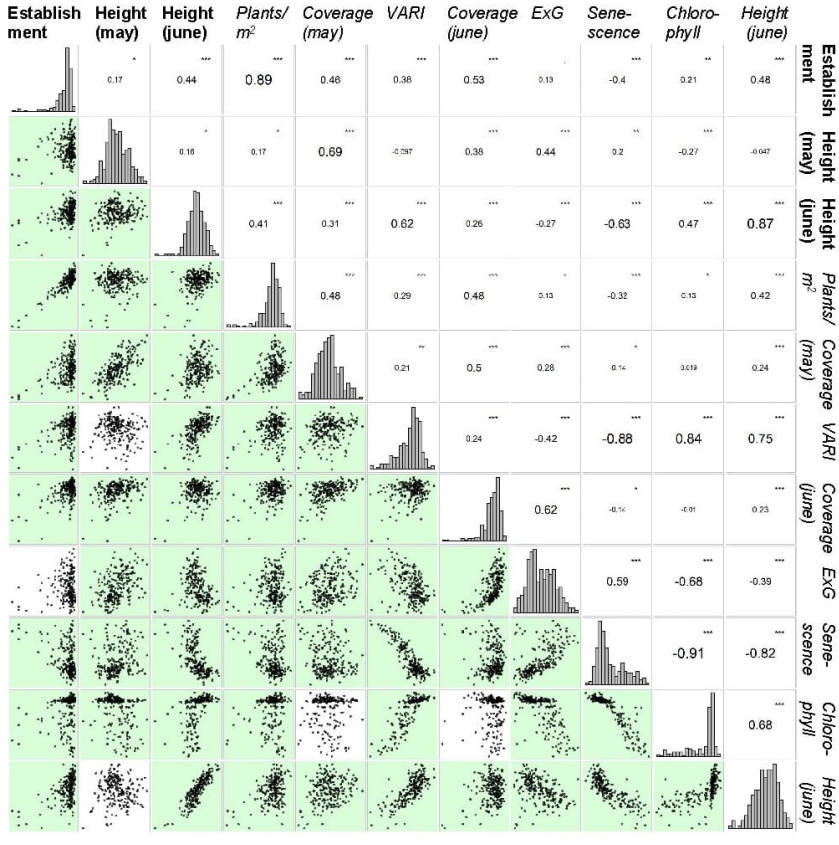


Figure 9 Correlation between drone assessments (in italics) and manual phenotyping (in bold). Analysed by: A. Chawade.

3.2.1 Earliness traits for Nordic adaption

Substantial phenotypic variation was observed across all agronomic and seed quality traits, highlighting the diversity within the panel and its relevance for pre-breeding. The following selected results are presented as mean values, based on measurements of individual accessions across years and replicates.

Since faba bean requires a relatively long period to reach pod maturity, earliness—the ability to mature within a shorter timeframe—is one of the most critical traits for successful cultivation in northern regions with short growing seasons. Key agronomic traits influencing earliness include flowering and maturation time. Early flowering is important as it ensures adequate time for pod filling before the end of the season, enables seed development during peak light periods, and helps mitigate risks from late-season frost or adverse weather conditions. Similarly, early maturity is crucial as it allows harvest to occur before autumn rains or frost, reducing drying costs, and facilitating timely sowing of winter crops in rotation systems. In our study, days to flowering and days to maturity varied by more than two weeks between early and late accessions of the diversity panel (Figure 10). Pairwise correlations of the phenotype data revealed that increased days to flowering correlated with increased plant height, a greater number of seeds per plant, and smaller seed size. As expected, field data also showed that later flowering correlated positively to days to maturity, reinforcing the strong relationship between those traits. Days to maturity showed a positive correlation with yield, suggesting that a longer maturation period contributes to higher seed production. Interestingly, no correlation between days to flowering and yield was observed. Notably, many early-maturing varieties in the diversity panel were non-commercial accessions, whereas later-maturing varieties were more frequently commercial cultivars. It is important to note, however, that the maturation times reported in this study may not reflect true harvest times. In practice, farmers often allow faba bean to fully dry in the field before harvesting, which extends the effective maturation period.

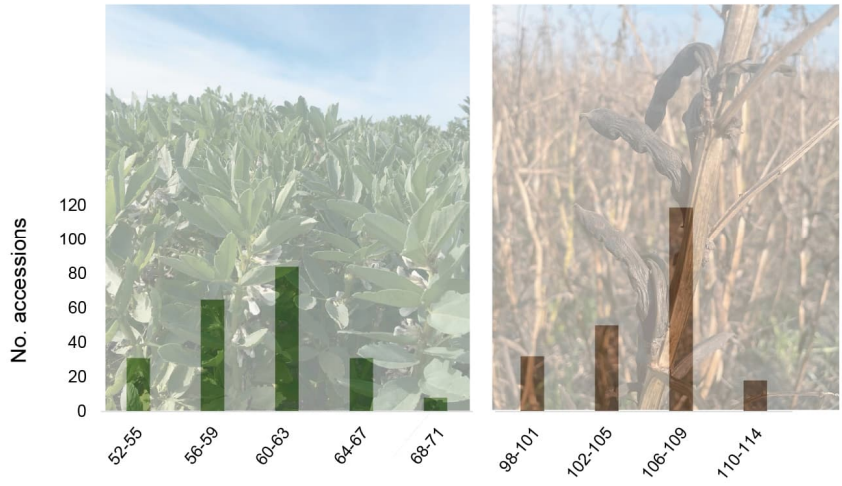


Figure 10 Field trial phenotype data for days to flowering (left) and days to maturity (right), showing the number of accessions (out of 220) that had reached flowering or maturity at each recorded time interval (days from sowing). Data represent the mean across two years and two replicates, illustrating the temporal distribution of these key agronomic traits within the diversity panel. Photos: H. Ohm & Å. Grimberg. Composition: E. Rémy.

3.2.2 Morphological diversity in plant and seed traits

In terms of flower colour, most accessions (192) displayed the typical variegated pattern, characterised by white petals with black spots and veins. Eight white-flowered varieties were specifically included for their low tannin content and therefore suitability for animal feed applications. A few accessions with brown and red flowers were also part of the panel (Figure 11A). Despite its potential relevance to the horticultural market, flower colour remains an underexplored trait in faba bean breeding, apart from its association with low-tannin white-flowered varieties (Hughes *et al.* 2020b). Varieties with red or purple flowers, particularly those with compact stature and profuse flowering in dwarf forms like *NGB22540/DWARF Ö53* (Figure 11B), may hold promise for expanding the appeal of faba bean into ornamental horticulture or home gardens, beyond traditional agricultural uses. Measurements of plant height revealed substantial diversity within the panel, with shorter varieties being nearly half the height of the tallest ones (Figure 11C). The dwarf variety predictably occupied the lowest end of the

height spectrum. During the greenhouse seed amplification process, a segregation pattern was observed in the variety *Ticol*, resulting in a tall and a short phenotype, that remained consistent across several generations.

Seed traits, including testa colours and seed size, represent some of the most diverse characteristics of the broad bean. Figure 11D illustrates the remarkable morphological variability observed in seed size and colour. Historically, selection in faba bean was likely based primarily on visual traits of seed morphology, influencing the classification of distinct subspecies by size—such as *minor*, *major*, *equina*, or *paucijuga* types. However, these classifications are now understood as varietal differences or botanical types resulting from extensive periods of human selection (Duc *et al.* 2010; Jayakodi *et al.* 2023). While this study aimed to incorporate a wide range of seed sizes in the diversity panel of faba bean accessions, it is important to acknowledge that breeding efforts have traditionally prioritized smaller seeds. This focus has been driven by the constraints of existing farming machinery, which is primarily designed for cereal crops but also adapted for e.g. peas.

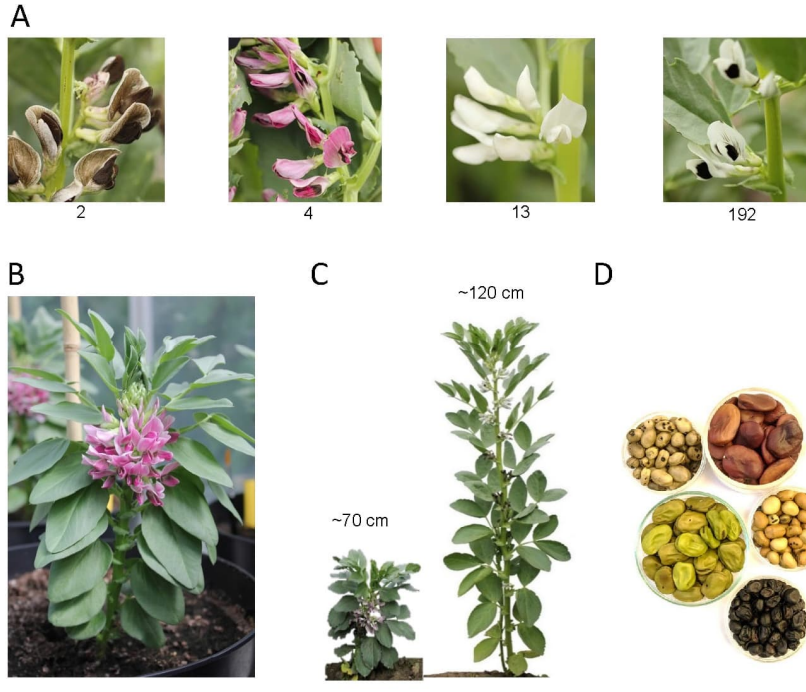


Figure 11 Morphological variation in faba bean diversity panel. A. Variation in flower colours observed within the diversity panel. The number of accessions in the diversity panel corresponding to each colour is indicated below the image. B. Compact growth habit of faba bean accession Ö53. C. Examples showcasing variation in plant height among different faba bean accessions. D. Seed size and seed colour variation in faba bean. Photos: H.Ohm, K. Baytes, A. Nieto-Esteve.

3.2.3 Trade-offs in breeding advancements

To assess how prior breeding progress has changed traits in faba bean, accessions were categorized based on registration status and breeding history (when available) and correlated these classifications with our phenotypic data. The results demonstrate significant differences in key agronomic traits, with accessions in the categories advanced or registered varieties exhibiting higher yield, seed number per plant, and plant height compared to the accessions from the categories landraces or cultivated. However, these advancements seem to have been accompanied by trade-offs**.** Delayed flowering and maturity were observed in the more advanced accessions, likely due to selection for prolonged vegetative growth enhancing seed production but also extending the growing cycle. This finding is particularly

relevant for Nordic breeding efforts**,** as it underscores the potential of selecting within the less advanced germplasm to find accession matching the need for earliness. While taller plants with extended vegetative phases generally produce more seeds, these traits can be problematic in short-season environments, where late-maturing varieties risk yield losses due to early autumn frosts or suboptimal drying conditions. The presence of early-flowering accessions with competitive yields in the diversity panel suggests a promising genetic reservoir providing breeding opportunities critical for adapting faba bean cultivation to short growing seasons. Further, breeding advancement has increased number of seeds while decreasing seed size, likely due to agronomic constraints linked to mechanical harvesting mentioned above.

3.3 Genome-wide association studies in faba bean

Advances in next-generation sequencing, particularly genotyping-by-sequencing (GBS) have increased marker density and reduced costs, making GWAS feasible. GWAS facilitates efficient selection and breeding strategies by identifying genetic variants associated with phenotyped traits, enabling the dissection of complex traits with higher resolution and improving the accuracy of marker-trait associations. Conventional breeding is slow, but genomic tools like marker-assisted selection (MAS) and quantitative trait loci (QTL) mapping improve precision by identifying regions of DNA linked to phenotypic traits, whether controlled by single major genes or multiple minor genes. GWAS has become an essential tool for uncovering the genetic architecture of complex traits in crops, offering higher resolution than traditional bi-parental QTL mapping by leveraging natural genetic variation in diverse populations. However, its power is influenced by factors such as population size and structure, and marker density (Kailash et al. 2017). While GWAS offers high-resolution mapping, it also has limitations, including potential false positives and data analysis complexity (Torkamaneh & Belzile 2022). Therefore, validation in independent populations remains essential before applying markers in breeding programs.

Recent GWAS studies in faba bean have identified single nucleotide polymorphisms (SNPs) linked to flowering time, seed size, and disease resistance (Faridi *et al.* 2021; Abou-Khater *et al.* 2022; Gulisano *et al.* 2023; Jayakodi *et al.* 2023; Karaköy *et al.* 2023; Skovbjerg *et al.* 2023; Gutierrez

et al. 2024; Lippolis *et al.* 2025). Existing genetic maps for faba bean require higher marker density for precise trait prediction, highlighting the need for improved mapping and validation efforts. Moreover, only a limited number of GWAS studies to date have utilized diversity panels.

3.3.1 Discovery of novel trait-associated markers

To expand genomic insights into faba bean, we conducted a GWAS on the diversity panel of 220 faba bean accessions to characterise genetic variation and identify trait-associated markers. Using phenotypic data from two years of field trials, we aimed to identify genetic markers linked to key agronomic and seed quality traits, providing valuable insights for breeding applications. In our study, we employed Diversity Arrays Technology Sequencing (DArTSeq), which claims to primarily target coding sequences (Szőke-Pázsi *et al.* 2024) integrating genome complexity reduction with high-throughput sequencing and has been effectively used in other legumes (Valdisser *et al.* 2017; Nkhata *et al.* 2020; Ahmed *et al.* 2021; Yohane *et al.* 2022; Lukanda *et al.* 2023). Despite resulting in a relatively high proportion of missing data (~30%), imputation improved the dataset, enabling the identification of 6,606 SNPs across the genome.

Among these, 51 novel SNPs were significantly associated with ten agronomic and seed quality traits, including thousand-grain weight, several seed size parameters, days to flowering, days to maturity, plant height, yield, and weevil susceptibility. Notably, 40 of these markers were located within predicted gene-coding regions of the reference genome, with several candidate genes implicated in key physiological processes. Furthermore, three SNPs associated with days to flowering were identified in predicted genes encoding proteins whose homologues in other plant species are known regulators of flowering.

The GWAS identified distinct markers for days to flowering and days to maturity with no overlap between them, suggesting these traits may be under independent genetic control. However, this apparent independence could also reflect low marker resolution, environmental effects masking potential relationships, or statistical artefacts. Interestingly, two flowering-associated markers were also linked to plant height and seed weight, suggesting potential pleiotropic effects where the same genetic regions influence multiple traits. This is further supported by our phenotypic observations, where increased days to flowering significantly correlated with increased

plant height and decreased seed weight. Expectedly, several markers for seed dimension traits overlapped, and one weevil susceptibility marker was also associated with seed weight and size. The identification of genetic markers associated with seed size parameters could offer new opportunities for targeted breeding, enabling selection beyond purely morphological traits. Many agronomic traits, such as flowering time, yield, and seed quality traits, are complex and governed by multiple genes with small to moderate effects (Falconer & Mackay 1996). This polygenic nature likely explains why some associations identified in this study were weaker or distributed across multiple genomic regions rather than concentrated in a single major-effect locus. Unlike previous GWAS studies of faba bean that focused on narrower genetic pools, our diversity panel captured a broader genetic base, increasing the likelihood of identifying novel associations. Given the species' large genome size and the reliance of previous studies on narrower genetic pools, the lack of overlap between our findings and previously reported markers is unsurprising.

Minor allele frequency (MAF), which denotes the frequency of the less common allele at a given locus within a population, is a crucial measure for evaluating genetic diversity. A low MAF indicates that one allele is rare in the population; while a MAF closer to 0.5 indicates more balanced frequencies between the two alleles, representing maximum genetic variation at that locus. To maximize marker discovery, a detection threshold of MAF = 0.01 was set, ensuring that no potentially informative markers were excluded. Although this relatively low threshold increases the risk of false positives, all significant SNP-trait associations found in our dataset, had MAF values exceeding 0.05, minimizing concerns regarding rare allele bias.

Population structure analysis using kinship and principal component analysis (PCA) confirmed limited relatedness between accessions and no major population structure. This is advantageous for GWAS, as it enhances the power to detect marker-trait associations without confounding due to shared ancestry. At the same time, the high genetic diversity within the population allows for a broad representation of phenotypic variations, a critical factor when studying complex traits that are influenced by multiple genetic backgrounds.

Additionally, the high broad-sense heritability (H^2) values observed for several of the 11 assessed traits suggest that phenotypic variation is largely driven by genetic factors, highlighting the potential for genomic selection to

enhance breeding precision for these traits. H^2 represents the proportion of total phenotypic variance attributed to genetic factors, including both additive effects (the cumulative impact of individual alleles) and non-additive effects (such as dominance and epistasis, where interactions between alleles influence the trait). However, as mentioned introductory, a diversity panel can generally make the heritability values less precise compared to more uniform backgrounds.

Linkage disequilibrium (LD) refers to the non-random association of alleles at different loci, where certain allele combinations occur more frequently than expected based on recombination rates and population history. LD influences the detection of SNP-trait associations in GWAS, as it determines whether markers remain linked to causal variants. Our LD decay revealed a rapid decline in LD with increasing genetic distance, suggesting frequent recombination events in the population's history. However, the observed low LD patterns should be interpreted with caution, as the large size of the faba bean genome combined with relatively low marker density means we may be undersampling LD patterns, potentially missing stronger associations in regions with sparse marker coverage.

3.3.2 Validation and breeding application of GWAS findings

To ensure the impact of these GWAS findings, further validation of associated markers and refinement are necessary. Replication of phenotypic data in field trials across independent locations, particularly in multi-environment trials, is crucial for confirming marker-trait associations under diverse growing conditions. In our study, SNPs were associated with traits of which several are generally sensitive to environmental variation. Multi-environment trials can thus help determine the stability of these associations and their applicability across different agricultural settings. Through such experiments, G×E interactions can be examined to determine whether genetic factors lead to varying phenotypic expressions under different environmental conditions, including microclimatic differences within field trials. While these interactions may reduce selection accuracy in breeding, they also offer opportunities to align specific traits with optimal environments, ultimately refining breeding strategies. A deeper understanding of GxE interactions enables breeders to stabilise yield across diverse environments and improve crop adaptability.

Integrating the results from our GWAS with other GWAS datasets through meta-analyses could enhance statistical power and refine genetic associations, allowing for more precise identification of trait-linked SNPs. Since genome-wide LD decay plots average LD across all chromosomes, they may mask localized high-LD regions. Generating LD heatmaps for significant GWAS regions thus allows for better distinction between causal and linked non-functional SNPs. Therefore, conducting further LD analysis would help validate the robustness of the GWAS and refine individual loci. By fine-mapping these regions, LD analysis can pinpoint causal variants, improving marker-assisted selection and increasing the accuracy of genomic predictions in breeding applications.

Furthermore, functional studies are essential to determine the roles of the identified significant SNP markers, particularly those linked to days to flowering, which are located in genes with identified homologs involved in flowering regulation. These genes are likely involved in flowering regulation, with a high probability of either directly influencing or targeting a functional gene. Investigating their underlying genetic mechanisms will determine their relevance as breeding targets, particularly the SNPs linked to flowering regulation, such as those associated with dolichol kinase, polyadenylate binding protein interacting protein, and chromatin-remodelling ATPase INO. This intermediate step, from marker discovery to functional validation, is crucial for distinguishing causative variants. Experimental approaches such as gene expression analysis or knockout studies can be employed to validate the biological relevance of these markers and their roles in trait expression.

4. From the flower to the seed

Improving agronomic productivity and nutritional quality in faba bean requires genetic and molecular insights into key developmental traits, particularly seed development and plant growth architecture (**Objectives II & III**). These interconnected traits influence yield potential, crop stability, and seed composition, making them critical targets for breeding and genetic improvement. As the mature seed represents the crop's primary edible and economic component, seed development is particularly significant, making an understanding of the seed-filling processes and nutrient partitioning essential. At the same time, flowering represents a pivotal phase in the plant's life cycle, marking the transition from vegetative growth to reproductive development which will result in seed formation. The timing and regulation of flowering influence plant architecture. The potential regulatory mechanisms behind determinate versus indeterminate growth in faba bean were investigated, which may affect branching patterns, flowering time, and harvest. Understanding these processes at a molecular level provides fundamental knowledge about the plant that might influence breeding targets for tailoring faba bean varieties. In **Papers I & III**, these developmental processes are examined in detail using a combination of transcriptomic, genetic, and biochemical methodologies. The following sections will explore these processes, starting with seed development and nutrient storage, followed by the identification of genes regulating flowering and growth architecture.

4.1 Seed development and nutrient storage

The seed serves as both a reproductive unit and a nutrient reservoir, surviving environmental stresses and remaining viable during storage and dispersal,

while providing the essential macromolecules required for seedling growth upon germination. The mature seeds' role as storage organs is the result of a fine-tuned developmental program, involving coordinated genetic, metabolic, and physiological activities so that nutrients are synthesised and compartmentalised efficiently (Gallardo *et al.* 2008; Gao *et al.* 2009). Metabolic biosynthetic pathways rely on the availability of carbon (from photosynthesis) and nitrogen (from soil uptake or symbiotic nitrogen fixation) and are essential for determining the final nutrient composition. The ratio of the major storage compounds starch, protein and oil varies significantly among plant species and even within the legume family (Song *et al.* 2017). For example, legumes such as soybean prioritise oil and protein synthesis, while others, like lentils, chickpeas or faba bean, specialise in accumulating protein and starch reserves (Table 1). The accumulation of seed storage proteins during seed development has been extensively studied in legumes (Golombek *et al.* 2001; Gallardo *et al.* 2008), including faba bean (De Pace *et al.* 1991; Panitz *et al.* 1995; Warsame *et al.* 2022). However, the underlying biological processes governing the storage of other macronutrients, such as starch and lipids, remain less well understood. While foundational research on faba bean seed development and seed filling has been conducted (Borisjuk *et al.* 1995, 2002), the regulatory networks controlling nutrient accumulation are still largely unexplored. Further investigations are crucial for advancing breeding strategies aimed at improving seed quality and yield (Kang *et al.* 2016).

To elucidate these processes, Paper I identified key developmental stages in faba bean seed filling, combining spatiotemporal transcriptome profiling and chemical nutrient analysis. Genes exhibit differential expression across various organs, tissues, or cells, as well as in response to stress or other adaptive mechanisms. These spatio-temporal patterns of gene expression are regulated by transcription factors, which serve as crucial regulatory proteins that control gene activity by binding to specific DNA sequences in response to various cellular signals. RNA sequencing (RNA-Seq) is a high-throughput sequencing method that enables transcriptome analysis by sequencing complementary DNA (cDNA) synthesized from RNA, allowing for the detection of both known and novel transcripts. In faba bean, RNA-Seq has been applied in various research areas, including gene characterisation, functional genomics, and gene expression analysis (Khan *et al.* 2019; Gao *et al.* 2020; Yang *et al.* 2020). This study quantified the accumulation of

protein, starch, and lipid content in both the embryo and endosperm across developmental stages and identified differentially expressed transcripts (DETs) associated with storage compound biosynthesis. Furthermore, as the first of its kind, this study provides insights into the transcriptional regulatory networks governing seed maturation and reserve accumulation. By analysing differentially expressed transcripts across developmental stages and seed tissues, this study aimed to improve the understanding of nutrient storage and regulation of seed development in faba bean. Transcripts investigated were homologous to genes in closely related species, including *Medicago truncatula* (barrel medic), *Pisum sativum* (pea), and *Glycine max* (soybean), annotated to encode key enzymes involved in metabolic pathways or transcription factors relevant to seed regulation. A simplified overview of methods and key findings is presented in the graphical abstract (Figure 12).

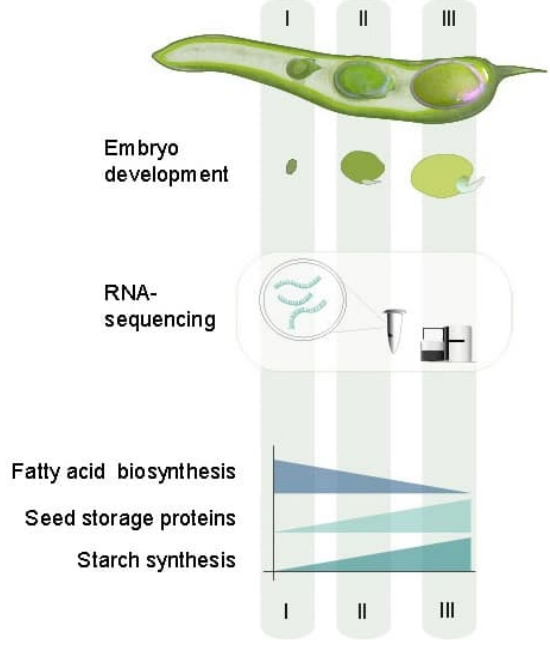


Figure 12 Simplified overview of Paper I, illustrating embryo development across three stages (I, II, III) and RNA sequencing for gene expression profiling during seed filling. Gradients represent key biosynthetic pathways: fatty acid biosynthesis (blue), seed storage protein synthesis (light blue), and starch synthesis (turquoise), highlighting their accumulation patterns. Illustration: H. Ohm & E. Rémy.

4.1.1 Stages of seed development and its regulators

Seed development in legumes progresses through three distinct phases: histodifferentiation, seed filling, and desiccation – and across three primary tissues: the embryo, endosperm, and seed coat (Borisjuk *et al.* 1995), see Figure 2 for tissue depiction. During histodifferentiation, the foundational structures of the seed are established. This phase can be divided into two sub-phases: the first involves cell division in the maternal and filial tissues, including the endosperm and seed coat, while the second focuses on the embryo, where rapid cell division shapes its basic architecture (Weber *et al.* 2005). As the seed transitions to the filling phase, cell enlargement is the major process, marked by the accumulation of storage compounds. The seeds' final maturation is in the desiccation phase, where seeds are prepared for storage and dispersal. Water content is reduced to about 10%, and the synthesis of protective molecules, such as late embryogenesis abundant proteins and sugars like raffinose and trehalose, ensures cellular stability during dehydration. This phase is critical for seed longevity and viability, as it equips seeds to withstand environmental fluctuations until conditions are favourable for germination.

In paper I, we defined four distinct stages of seed development in faba bean, spanning approximately 15 to 41 days after flowering. During this period, the seed size, embryo-to-endosperm ratio and dry matter content steadily increased. These stages outline the sequential progression from early embryo expansion to the later phases of storage compound accumulation and seed desiccation, providing a framework for the developmental timeline and the basis for further metabolic and transcriptomic analysis. DETs encoding transcription factors that in other plant species are known to regulate seed development were mapped. This included the identification of transcripts homologous to the LAFL transcription factor network (LEAFY COTYLEDON 1 (LEC1), ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3), and LEC2), master regulators of storage compound synthesis and seed maturation (Gazzarrini & Song 2024). Our data align with previous findings suggesting that *LEC1* acts as a pioneering master transcription factor during early seed development, while *FUS3* and *ABI3* drive maturation-specific processes in later stages (Jo *et al.* 2019). However, DETs homologous to *LEC2* were absent in our dataset as well as in the ORFs of reference genome Hedin/2 (Jayakodi et al., 2023). Additionally, the LAFL networks' major repressors, homologs of *VAL1* and *ASIL1* (Jia *et al.* 2014),

were also identified as DETs in our dataset, highlighting their regulatory roles in seed development. Notably, genes encoding transcription factors are promising targets for breeding, as they can exert significant influence on phenotypic traits.

4.1.2 Lipid accumulation

In plants, fatty acids are initially synthesized in the chloroplasts and subsequently transported to the endoplasmic reticulum as acyl-CoA, where they are esterified with glycerol to form various lipid species, including triacylglycerols (TAGs), which are stored in oil bodies (Nikiforidis 2019). As key intermediates in lipid metabolism, free fatty acids (FFAs) contribute to TAG synthesis and can be mobilised through lipolysis when energy is required. Lipid content in the developing faba bean seed, represented by FFAs and TAGs, was analysed using thin-layer chromatography to separate lipid classes, followed by methylation into fatty acid methyl esters and subsequent quantification via gas chromatography. Results demonstrated a significant increase in TAG levels throughout embryo development, with concentrations rising fivefold between early and late stages, while endosperm TAG levels remained relatively stable. In contrast, FFA levels were significantly lower (approximately one-tenth) compared to TAG and were lower in the embryo than in the endosperm tissue. In the embryo, FFA levels exhibited a declining trend towards later stages of development, likely due to their rapid utilisation in TAG biosynthesis or membrane formation. The opposite was observed in the endosperm.

Nearly all major enzymatic steps of the fatty acid biosynthesis pathway (occurring in the chloroplasts) exhibited decreasing transcript expression during embryo development. A similar decline was observed in transcripts encoding enzymes catalyzing the majority of steps from acyl-CoA synthesis to TAG formation (occurring in the endoplasmic reticulum), as well as those involved in TAG breakdown into fatty acids and glycerol. Transcripts associated with beta-oxidation, converting FFAs into acetyl-CoA in peroxisomes for energy production, tended to be enriched in later embryo development stages (but showed inconsistent trends between the two varieties analysed). Similarly, transcripts encoding plant oil body proteins, such as oleosins, which are closely associated with TAG storage and play a key role in stabilizing oil bodies during seed maturation, showed an opposite trajectory, increasing at later developmental stages, potentially reflecting

their role in stabilizing lipid storage structures as seed maturation progresses.

Interestingly, we observed a decline in the expression of a transcript homologous to *WRINKLED1* (*WRI1)*, a gene encoding a transcription factor that promotes oil synthesis during embryo maturation by activating genes involved in glycolysis and fatty acid biosynthesis (Ma *et al.* 2013). The early peak in *WRI1* expression during embryo development, followed by the delayed increase in oil body proteins, suggests that while faba bean has the genetic framework for oil accumulation, the regulatory balance favours metabolic turnover over storage. While this gene expression pattern in faba bean bears similarity to that of high oil-accumulating seeds, the lack of simultaneous high expression of both *WRI1* and oleosin likely contributes to the relatively low final oil levels in mature seeds. This implies that in the absence of well-organized lipid droplets, fatty acids and TAGs may be more readily turned over rather than effectively stored. This pattern is consistent with the increased expression of genes involved in beta-oxidation, which could indicate that fatty acids are being mobilized for energy production rather than stored as long-term reserves.

4.1.3 Starch accumulation

Starch is the predominant carbohydrate in mature faba bean seeds, formed from photosynthates transported as sucrose through the phloem (Borisjuk *et al.* 2002). These sugars are either used immediately for growth, temporarily stored as starch, or allocated to amyloplasts, specialised organelles responsible for starch storage, for long-term reserve accumulation (MacNeill *et al.* 2017). We determined the total starch content of the developing faba bean enzymatically by hydrolysing starch with α-amylase and amyloglucosidase, followed by colourimetric quantification of the resulting glucose. Results showed that starch accumulation in the embryo increased progressively during seed development, whereas endosperm starch levels remained low and relatively stable. Similarly, transcripts related to starch and sucrose metabolism generally showed increased expression in the embryo throughout seed development. Histochemical staining of faba bean seed sections with Lugol's iodine revealed a starch-rich section in the seed coat during early development, suggesting that these structures store transient starch that serves as a temporary carbon source for the maturing embryo, as has been seen in pea (Quilichini *et al.* 2022).

4.1.4 Protein accumulation

Seed storage proteins, despite structural differences across plant species, share common features of accumulating in specific seed tissues during distinct developmental stages, predominantly regulated at the transcriptional level (Verdier *et al.* 2008). Protein content was quantified using the Dumas method, which measures total nitrogen via elemental analysis and applies a conversion factor of 6.25 to estimate protein levels. Our results showed that total protein content in faba bean embryos was already high at early developmental stages and declined slightly toward seed maturation. Protein levels were slightly higher in the endosperm compared to the embryo. This trend of declining protein levels during seed development aligns with previous findings in faba bean and other legume species (Sital *et al.* 2011; Lu *et al.* 2016; Zhang *et al.* 2021; Warsame *et al.* 2022). Seed storage protein transcripts, including those homologues to genes encoding cupin, legumin, globulin, and vicilin, were highly differentially expressed in seed tissues, with expression increasing during embryo development.

4.1.5 Differences between the embryo and the endosperm

In the dicotyledonous faba bean, seed development is characterized by a growing embryo at the expense of a decreasing endosperm tissue. Except for protein content, the endosperm consistently exhibited lower nutrient levels than the embryo, with similar accumulation patterns observed across both studied varieties. The embryo tissue exhibited more DETs related to transcription factors during seed development than the endosperm, suggesting more complex developmental and metabolic shifts. This aligns with observed differences in storage compound accumulation, where starch and oil levels increased in the embryo but remained constant in the endosperm. Interestingly, protein levels stayed high and relatively stable in both tissues during the first three stages of seed development. These insights highlight the complex regulatory systems that control storage compound accumulation in seeds, and the developmental cues influencing these processes.

4.2 Flowering and the *PEBP* family structure in faba bean

A key aspect of understanding agriculturally important traits (**Objective III**) is uncovering the genetic and molecular mechanisms that regulate flowering. Various factors, including photoperiod, temperature, and hormonal pathways tightly regulate this process. Among the core gene families controlling flowering, the one encoding phosphatidylethanolamine-binding proteins (PEBPs) is of particular importance. The *PEBP* gene family encompasses diverse functions that regulate flowering time, plant architecture, and other developmental processes, with members of this family acting as either promoters or inhibitors of flowering, depending on environmental context and species (Karlgren *et al.* 2011). In both model species, such as Arabidopsis, and economically important crops, including legumes, *PEBPs* play a central role in flowering regulation and subsequent pod development, directly impacting crop productivity (Su *et al.* 2024). While flowering time is regulated by multiple gene families including *MADS-box* and *CONSTANS*, *PEBP*-genes are particularly significant due to their conserved role as mobile flowering signals and their demonstrated importance in legume crop adaptation, making them prime targets for understanding and improving flowering control in faba bean (Kong *et al.* 2010; Liu *et al.* 2018; Williams *et al.* 2022). Consequently, a detailed study of *PEBP* genes in faba bean provides critical insights into the genetic and physiological basis of flowering, offering potential avenues for enhancing crop yield and adaptation to diverse environments.

In paper III, we identified 11 Vf*PEBP* genes in *Vicia faba* for the first time. Phylogenetic analysis classified them into MFT-like, TFL-like, and FT-like subfamilies, revealing their relationship to *PEBP* genes from closely related legumes *Pisum sativum* and *Vicia sativa*. A comparison of gene numbers within these subfamilies across several legume species suggests a more specialised and complex functional pattern relative to *Arabidopsis*. For example, a single *FT* gene in *Arabidopsis* regulates the transition from vegetative to reproductive growth. However, in legumes such as *Medicago truncatula* and *Glycine max*, *FT* has expanded into six and twelve paralogs respectively, through both ancient duplications leading to functional diversification and, in soybean, a more recent whole genome duplication event. These paralogs have acquired distinct functional roles - for instance, *GmFT1a* acts as a floral repressor under long-day conditions, whereas

GmFT2a has the opposite role, highlighting the functional divergence of this gene family in legumes (Liu *et al.* 2018).

4.2.1 Genetic basis of determinacy

Plant architecture, particularly the balance between determinate and indeterminate growth, is a key agronomic trait influencing reproductive timing and spatial structure. Most faba bean varieties exhibit indeterminate growth (Figure 13A), continuing vegetative development and flowering until senescence. In contrast, determinate varieties (Figure 13B), transition early to a terminal inflorescence, ceasing vegetative growth and resulting in a shorter flowering period. Determinate growth could lead to a more synchronized ripening and denser plant architecture, which can offer advantages in agricultural settings.

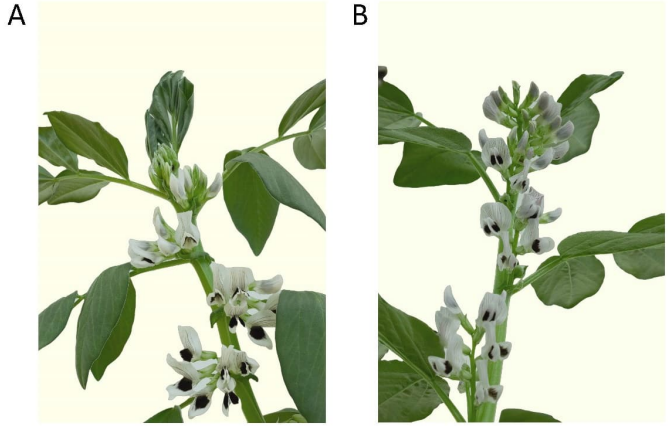


Figure 13 Indeterminate (A) and determinate (B) variations of faba bean. Photo A. Nieto-Esteve.

In faba bean, the *PEBP* member *Terminal Flower1* (*VfTFL1*) gene is suggested to be a key regulator of this process by maintaining an indeterminate inflorescence and delaying floral meristem transition. A single SNP within the first exon of the coding region has been proposed to lead to the functional mutation of *VfTFL1* as a potential marker for differentiating between indeterminate and determinate faba bean varieties (Avila *et al.* 2006, 2007). In pea, two paralogs of *TFL1*, *PsDET* and *PsLF*, have been identified as one delaying floral transition and the other sustaining

inflorescence indeterminacy (Foucher *et al.* 2003; Benlloch *et al.* 2015). Our phylogenetic analysis of *PEBP* genes indicates that these correspond to *VfTFL1a* and *VfTFL1c* in faba bean, and an analysis of their putative cis-regulatory elements revealed that they are predicted to be predominantly regulated by pathways associated with light, hormonal signalling, and abiotic stress responses. By amplifying, cloning, and sequencing *VfTFL1a* from four indeterminate and four determinate varieties, we identified allelic variations within the gene, as well as in the flanking upstream and downstream regions. None of the investigated genetic variants, including the previously proposed marker, could distinguish determinate from indeterminate faba bean varieties. This contradicts earlier findings on a genetic marker for predicting determinacy, highlighting the need for further research to clarify the genetic mechanisms governing this trait. Phenotypic data from two years of field trials revealed that determinate types flowered significantly later and exhibited reduced variability than the indeterminate types across all assessed traits, including plant height, days to flowering, days to maturity, yield, seeds per plant, and seed size. This suggests that determinate growth confers more uniform phenotypes. While this apparent stability could partly reflect genetic relationships within the determinate group, the consistent pattern across multiple traits suggests a biological basis worth further investigation. However, the observed delay in flowering and maturation indicates these determinate varieties may be less suitable for Nordic regions, where early maturity is a key breeding target.

5. Methodological considerations

Several technical challenges encountered in the work described in this thesis highlight critical areas for methodological improvement for further faba bean research.

One of the methodological constraints in this study was the relatively low seed amplification rate of faba bean, particularly in comparison to crops such as cereals and oilseeds. Gene banks typically provide only a limited number of seeds per accession, with standard distributions often as low as 30 seeds per genotype. This necessitates an initial seed multiplication phase before the commencement of large-scale field trials. Each genotype was expected to yield between 300 and 1.500 seeds from the isolated plants in our amplification—a quantity deemed sufficient to sustain field trials across two locations over two consecutive years. However, early spring drought and high rates of bird predation during germination led to significant plant losses, reducing the available seed stock for subsequent experiments. Additionally, accessions with exceptionally large seeds were often incompatible with standard mechanical sowing, requiring manual planting or, in some cases, exclusion from the trials to ensure uniform sowing conditions. To mitigate this, additional seed amplification was conducted in greenhouse conditions for selected accessions.

Spring-sown faba bean, such as in our field trials, rely heavily on early spring and summer precipitation, requiring early sowing for optimal yields while being vulnerable to summer drought (Jensen *et al.* 2010). This vulnerability was evident in the field trials, where the 2021 trial required irrigation in the early season and also faced a late-summer drought. The differences in precipitation patterns resulted in noticeable phenotypic differences between years, for example with increased plant height and yield

observed in 2022, a year characterised by more evenly distributed rainfall throughout the growing season.

In our GWAS, an initial GBS approach based on leaf sampling in the field failed due to high levels of bacterial DNA contamination (*Pantoea agglomerans*). It remains unclear whether this contamination arose due to external environmental factors (e.g., an unusually high microbial load) or suboptimal restriction enzyme (RE) selection during library preparation. To overcome these shortcomings, we used greenhouse-grown plants and employed the GBS technique DArTSeq as well as an alternative sampling method to minimize the risk of bacterial contamination. A limitation of our pooled leaf sampling method (five individuals per accession) was the challenge of obtaining allele frequency data since the binary SNP format from the DArTseq dataset does not allow for a differentiation between homozygous and heterozygous states of individuals. As a result, individual allele frequencies could not be reconstructed, limiting the ability to quantify genetic diversity, assess population structure, or detect rare alleles relevant for breeding. Future research could improve allele frequency estimation by sequencing individuals separately or using sequencing methods tailored for pooled samples, or optimise the GBS protocol, such as the restriction enzyme (RE) comparisons conducted recently by Zhang *et al.* (2024), to enhance accuracy in genotyping complex plant genomes.

RNA extraction from plant tissues presents notable technical challenges, particularly in achieving RNA of sufficient integrity and quality for downstream applications. In this study, these difficulties were likely compounded by the presence of secondary metabolites in the plant tissues, which are known to interfere with RNA extraction and stabilisation. As a result, two RNA samples from the full set of samples used in this study failed to meet the required quality threshold for sequencing and were therefore excluded from further data processing. Further, technical difficulties with the physical separation of developing seed tissues, particularly that of endosperm from the inner seed coat at a late developmental stage, may result in overlapping transcriptional profiles masking tissue-specific patterns. Future studies would benefit from refined microdissection techniques or single-cell transcriptomics to resolve tissue-specific expression dynamics.

The amplification of specific DNA fragments in faba bean was complicated by its repetitive genome structure, with repeated segments extending several hundred base pairs. This genomic complexity, combined

with high intervarietal variation and strand switching during amplification, necessitates the development of more robust amplification strategies. The presence of secondary metabolites in leaf tissue further affected DNA purity and PCR efficiency, suggesting the need for optimised extraction protocols.

Regarding nutrient profiling, methodology choice can affect the results, for example with regard to the absolute levels. Protein quantification using the Dumas method revealed potential overestimation due to non-protein nitrogen compounds in immature seeds, and due to the commonly applied conversion factor of 6.25 may be too high, leading to additional overestimation (Mossé 1990; Krul 2019). Regarding starch, it can be noted that the enzymatic method used in our study tends to yield slightly lower values compared to polarimetric approaches (Jezierny *et al.* 2017). However, in the context of this thesis, the focus is not on the absolute compound levels but rather on the comparative analysis of varietal differences and spatiotemporal variation.

6. Conclusion

This thesis provides new insights into faba bean genetics and development, integrating field-based, genomic, and transcriptomic approaches to address key breeding challenges. The characterisation of a global diversity panel under Nordic field conditions revealed substantial phenotypic variation, particularly in traits relevant to yield, earliness, and seed quality. This diversity not only highlights the potential of the germplasm for developing new elite breeding lines but also laid the foundation for a GWAS, which identified 51 novel SNPs linked to agronomic and seed quality traits. Of these, 40 were located in predicted gene-coding regions, providing a direct link to functional pathways and immediate targets for marker-assisted selection. Further, spatiotemporal nutrient and transcriptomic profiling of developing seed tissues mapped the regulatory networks underlying storage compound accumulation. Identification of gene expression patterns of stage-specific transcription factors, including members of the LAFL network, demonstrated their role in coordinating protein, starch, and lipid biosynthesis. The characterisation of the *PEBP* gene family provided the first comprehensive overview of flowering regulators in faba bean and challenged a previously presented diagnostic marker for determinacy. Plants with determinate growth habit exhibited significantly delayed flowering in field, while also displaying lower phenotypic variability compared to indeterminate varieties.

6.1 Implications and future directions

Earliness is a critical trait for successful faba bean cultivation in the short growing seasons of the Nordic region, with flowering time and maturity serving as key determinants of crop adaptability. In this study, genetic

markers were identified for both traits, yet no overlap was observed despite their strong phenotypic correlation. This suggests a degree of genetic independence between flowering time and maturity, a finding with implications for breeding strategies in Nordic environments. The potential to manipulate these traits separately offers opportunities to develop early-flowering cultivars without compromising yield, a particularly relevant goal given the observed trend that breeding advances have favoured later-flowering and later-maturing cultivars. Notably, early-maturing varieties were predominantly found among non-commercial accessions, highlighting valuable but underutilised genetic resources for early maturation in modern breeding programs. The identification of the *PEBP* gene family in faba bean provides further insights into flowering regulation, with particular interest in *VfTFL1* paralogues, given their roles in controlling both flowering time and plant architecture. Interestingly, while determinate varieties exhibited delayed flowering, they also displayed greater phenotypic stability. However, a detailed analysis of *VfTFL1a* did not reveal any genetic variation that distinguishes determinate from indeterminate varieties, indicating that additional regulatory mechanisms influence this trait. These could involve upstream or downstream regions beyond the ones investigated in the study, functional variation in other *VfTFL1* paralogues, or the involvement of transcription factors such as *FD*, the *SEP* genes, or other MADS-box proteins known to regulate determinacy in other species (Hugouvieux *et al.* 2018; Shah *et al.* 2022). Collectively, these findings highlight an untapped potential within faba bean germplasm to develop cultivars that combine early maturation with agronomic performance, warranting further investigation in breeding programs aimed at short-season environments.

Food quality traits of faba bean seeds represent critical breeding targets. Our findings provide a foundation for improving seed quality through breeding, offering new avenues to enhance the nutritional value of faba bean. Field trials indicate that increased protein content in faba bean does not necessarily compromise yield, challenging the commonly observed trade-off in other crops, such as soybean (Guo *et al.* 2022), where higher protein content is often linked to reduced yield due to carbon allocation constraints. Although further large-scale trials are required to confirm this, our results highlight significant opportunities for breeding programs aimed at enhancing nutritional value without sacrificing agronomic performance.

Transcriptomic analyses of seed development identified key regulatory networks governing storage compound accumulation, revealing several potential breeding strategies, for instance the identified LAFL network regulators of storage protein synthesis, could serve as targets for precise genetic modification through CRISPR/Cas9—although this approach remains in its early stages for faba bean. Additionally, the early expression of *WRI1* during seed development, followed by upregulation of oil body proteins and increased TAG levels, suggests a latent capacity for enhanced lipid biosynthesis in faba bean seeds. Investigating this mechanism experimentally could offer valuable insights into the constraints on oil accumulation in this species, indicating a potential breeding target for increasing oil content. This could enable the development of faba bean varieties with dual-purpose seed use, akin to soybean, where both protein and oil serve as valuable components. Finally, the diversity panel exhibited substantial variation in protein content, highlighting the potential of natural genetic diversity for breeding efforts. Comparative transcriptomic analyses of accessions with extreme protein levels—both high and low—could further elucidate the regulatory pathways governing protein biosynthesis, degradation, and storage dynamics.

References

Abbes, Z., Kharrat, M., Delavault, P., Simier, P. & Chaïbi, W. (2007). Field evaluation of the resistance of some faba bean (Vicia faba L.) genotypes to the parasitic weed Orobanche foetida Poiret. *Crop Protection*, vol. 26 (12), pp. 1777–1784 Elsevier. DOI: https://doi.org/10.1016/J.CROPRO.2007.03.012

Abou-Khater, L., Maalouf, F., Jighly, A., Alsamman, A.M., Rubiales, D., Rispail, N., Hu, J., Ma, Y., Balech, R., Hamwieh, A., Baum, M. & Kumar, S. (2022). Genomic regions associated with herbicide tolerance in a worldwide faba bean (Vicia faba L.) collection. *Scientific Reports*, vol. 12 (1). DOI: https://doi.org/10.1038/s41598-021-03861-0

Adhikari, K.N., Khazaei, H., Ghaouti, L., Maalouf, F., Vandenberg, A., Link, W. & O'Sullivan, D.M. (2021). Conventional and Molecular Breeding Tools for Accelerating Genetic Gain in Faba Bean (Vicia Faba L.). *Frontiers in Plant Science*, vol. 12, p. 744259 Frontiers Media S.A. DOI: https://doi.org/10.3389/FPLS.2021.744259/PDF

Aguilar-Benitez, D., Casimiro-Soriguer, I., Maalouf, F. & Torres, A.M. (2021). Linkage mapping and QTL analysis of flowering time in faba bean. *Scientific Reports 2021 11:1*, vol. 11 (1), pp. 1–11 Nature Publishing Group. DOI: https://doi.org/10.1038/s41598-021-92680-4

Ahmed, S.M., Alsamman, A.M., Jighly, A., Mubarak, M.H., Al-Shamaa, K., Istanbuli, T., Momtaz, O.A., El Allali, A. & Hamwieh, A. (2021). Genome-wide association analysis of chickpea germplasms differing for salinity tolerance based on DArTseq markers. *PLOS ONE*, vol. 16 (12), p. e0260709 Public Library of Science. DOI: https://doi.org/10.1371/JOURNAL.PONE.0260709

Akinwale, R.O., Odunlami, L.K., Eze, C.E. & Oladejo, A.S. (2021). Effectiveness of different alpha lattice designs in the evaluation of maize (Zea mays L.) genotypes in a rainforest agro-ecology. *Heliyon*, vol. 7 (7), p. e07414. DOI: https://doi.org/10.1016/j.heliyon.2021.e07414

Allier, A., Teyssèdre, S., Lehermeier, C., Moreau, L. & Charcosset, A. (2020). Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics*, vol. 21 (1), p. 349 BioMed Central Ltd. DOI: https://doi.org/10.1186/S12864-020-6756-0

Auvinen, P., Chang, W., Holm, L., James, M., Jääskeläinen, M., Khazaei, H., Laine, P.K., Paulin, L., Salgado, M., Frederick, L., Tanskanen, J., Törönen, P. & Schulman, A.H. (2023). A faba bean pan-genome for advancing sustainable protein security. (24), pp. 7–9

Avila, C.M., Atienza, S.G., Moreno, M.T. & Torres, A.M. (2007). Development of

a new diagnostic marker for growth habit selection in faba bean (Vicia faba L.) breeding. *Theoretical and Applied Genetics*, vol. 115 (8), pp. 1075–1082. DOI: https://doi.org/10.1007/s00122-007-0633-y

Avila, C.M., Nadal, S., Moreno, M.T. & Torres, A.M. (2006). Development of a simple PCR-based marker for the determination of growth habit in Vicia faba L. using a candidate gene approach. *Molecular Breeding*, vol. 17 (3), pp. 185–190. DOI: https://doi.org/10.1007/s11032-005-4075-4

Ávila, C.M., Ruiz-Rodríguez, M.D., Cruz-Izquierdo, S., Atienza, S.G., Cubero, J.I. & Torres, A.M. (2017). Identification of plant architecture and yield-related QTL in Vicia faba L. *Molecular Breeding*, vol. 37 (7)

Ballard, A. (2020). Our future Proteins. *IPPR Progressive Review*, vol. 27 (3), pp. 316–320

Bassi, F.M., Sanchez-Garcia, M. & Ortiz, R. (2024). What plant breeding may (and may not) look like in 2050? *The Plant Genome*, vol. 17 (1), p. e20368 John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/TPG2.20368

Benlloch, R., Berbel, A., Ali, L., Gohari, G., Millán, T. & Madueño, F. (2015). Genetic control of inflorescence architecture in legumes. *Frontiers in Plant Science*, vol. 6, pp. 1–14

Bernal-Gallardo, J.J. & de Folter, S. (2024). Plant genome information facilitates plant functional genomics. *Planta*, vol. 259 (5), p. 117 Springer Science and Business Media Deutschland GmbH. DOI: https://doi.org/10.1007/S00425-024-04397-Z

Bishop, J. & Nakagawa, S. (2021). Quantifying crop pollinator dependence and its heterogeneity using multi-level meta-analysis. *Journal of Applied Ecology*, Blackwell Publishing Ltd.

Björnsdotter, E., Nadzieja, M., Chang, W., Escobar-Herrera, L., Mancinotti, D., Angra, D., Xia, X., Tacke, R., Khazaei, H., Crocoll, C., Vandenberg, A., Link, W., Stoddard, F.L., O'Sullivan, D.M., Stougaard, J., Schulman, A.H., Andersen, S.U. & Geu-Flores, F. (2021). VC1 catalyses a key step in the biosynthesis of vicine in faba bean. *Nature Plants 2021 7:7*, vol. 7 (7), pp. 923–931 Nature Publishing Group. DOI: https://doi.org/10.1038/s41477-021-00950-w

Borisjuk, L., Walenta, S., Rolletschek, H., Mueller-Klieser, W., Wobus, U. & Weber, H. (2002). Spatial analysis of plant metabolism: Sucrose imaging within Vicia faba cotyledons reveals specific developmental patterns. *The Plant Journal*, vol. 29 (4), pp. 521–530. DOI: https://doi.org/10.1046/j.1365-313x.2002.01222.x

Borisjuk, L., Weber, H., Panitz, R., Manteuffel, R. & Wobus, U. (1995). Embryogenesis of Vicia faba L.: Histodifferentiation in Relation to Starch and Storage Protein Synthesis. *Journal of Plant Physiology*, vol. 147 (2), pp. 203–218

Brünjes, L. & Link, W. (2021). Paternal outcrossing success differs among faba bean genotypes and impacts breeding of synthetic cultivars. vol. 134, pp. 2411–2427. DOI: https://doi.org/10.1007/s00122-021-03832-z

Caracuta, V., Weinstein-Evron, M., Kaufman, D., Yeshurun, R., Silvent, J. & Boaretto, E. (2016). 14,000-year-old seeds indicate the Levantine origin of the lost progenitor of faba bean. *Scientific Reports*, vol. 6 Nature Publishing Group.

Chetto, O., Belqadi, L., El, Z., Fatemi, A., Charafi, J., Kouighat, M., Najmi, A., El, M., Houmanat, K. & Nabloussi, A. (2024). Assessment of elite faba bean lines for enhanced productivity and resilience in contrasting challenging environments using phenotypic and molecular markers. vol. 18 (September) Elsevier B.V.

Cooper, J.W., Wilson, M.H., Derks, M.F.L., Smit, S., Kunert, K.J., Cullis, C. & Foyer, C.H. (2017). Enhancing faba bean (Vicia faba L.) genome resources. *Journal of experimental botany*, vol. 68 (8), pp. 1941–1953 Oxford University Press. DOI: https://doi.org/10.1093/jxb/erx117

Crépon, K., Marget, P., Peyronnet, C., Carrouée, B., Arese, P. & Duc, G. (2010). Nutritional value of faba bean (Vicia faba L.) seeds for feed and food. *Field Crops Research*, vol. 115 (3), pp. 329–339. DOI: https://doi.org/10.1016/j.fcr.2009.09.016

Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F.N. & Leip, A. (2021). Food systems are responsible for a third of global anthropogenic GHG emissions. *Nature Food 2021 2:3*, vol. 2 (3), pp. 198–209 Nature Publishing Group. DOI: https://doi.org/10.1038/s43016-021-00225-9

Cruz-Izquierdo, S., Avila, C.M., Satovic, Z., Palomino, C., Gutierrez, N., Ellwood, S.R., Phan, H.T.T., Cubero, J.I. & Torres, A.M. (2012). Comparative genomics to bridge Vicia faba with model and closely-related legume species: Stability of QTLs for flowering and yield-related traits. *Theoretical and Applied Genetics*, vol. 125 (8), pp. 1767–1782

Duc, G., Bao, S., Baum, M., Redden, B., Sadiki, M., Suso, M.J., Vishniakova, M. & Zong, X. (2010). Diversity maintenance and use of Vicia faba L. genetic resources. *Field Crops Research*, vol. 115 (3), pp. 270–278

Duluins, O. & Baret, P.V. (2024). A systematic review of the definitions, narratives and paths forwards for a protein transition in high-income countries. *Nature Food 2024 5:1*, vol. 5 (1), pp. 28–36 Nature Publishing Group. DOI: https://doi.org/10.1038/s43016-023-00906-7

EC (2018). *Report from the commission to the council and the european parlaiament on the development of plant proteins in the European Union*. Brussels. Available at: https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/plants_and_plant_products/documents/report-plant-proteins-com2018-757-final_en.pdf [2020-01-03]

EC (2020). *EU agricultural outlook for markets, income and environment, 2020-2030*. Brussels. DOI: https://doi.org/10.2762/252413

EC (2024). *Protein Supply and Demand* Available at: https://agriculture.ec.europa.eu/farming/crop-productions-and-plant-based-products/cereals/reducing-plan-protein-deficit-eu_en

Ellwood, S.R., Phan, H.T.T., Jordan, M., Hane, J., Torres, A.M., Avila, C.M., Cruz-

Izquierdo, S. & Oliver, R.P. (2008). Construction of a comparative genetic map in faba bean (Vicia faba L.); conservation of genome structure with Lens culinaris. *BMC genomics*, vol. 9, p. 380 BioMed Central. DOI: https://doi.org/10.1186/1471-2164-9-380

Falconer, D.. & Mackay, T.F.C. (1996). *Introduction to Quantitative Genetics*. 4th. ed. Addison Wesley Longman, Harlow.

FAO (2019). *The State of the World's Biodiversity for Food and Agriculture, FAO Commission on Genetic Resources for Food and Agriculture Assessments.* Rome. Available at: http://www.fao.org/3/CA3129EN/CA3129EN.pdf [2020-08-04]

FAO (2024). *Agricultural Production - Our World in Data*. *UN Food and Agriculture Organization (FAO) – processed by Our World in Data.* Available at: https://ourworldindata.org/agricultural-production#legumes-and-nuts [2025-01-27]

FAO (2025). *FAOSTAT*. Available at: https://www.fao.org/faostat/en/#data [2025-02-03]

Faridi, R., Koopman, B., Schierholt, A., Ali, M.B., Apel, S. & Link, W. (2021). Genetic study of the resistance of faba bean (Vicia faba) against the fungus Ascochyta fabae through a genome-wide association analysis. *Plant Breeding*, vol. 140 (3), pp. 442–452 John Wiley & Sons, Ltd. DOI: https://doi.org/10.1111/PBR.12918

Ferreira, H., Vasconcelos, M., Gil, A.M. & Pinto, E. (2021). Benefits of pulse consumption on metabolism and health: A systematic review of randomized controlled trials. *Critical Reviews in Food Science and Nutrition*, vol. 61 (1), pp. 85–96 Taylor & Francis. DOI: https://doi.org/10.1080/10408398.2020.1716680

Flores, F., Nadal, S., Solis, I., Winkler, J., Sass, O., Stoddard, F.L., Link, W., Raffiot, B., Muel, F. & Rubiales, D. (2012). Faba bean adaptation to autumn sowing under European climates. *Agronomy for Sustainable Development*, vol. 32 (3), pp. 727–734 Springer. DOI: https://doi.org/10.1007/S13593-012-0082-0/TABLES/5

Foucher, F., Morin, J., Courtiade, J., Cadioux, S., Ellis, N., Banfield, M.J. & Rameau, C. (2003). Determinate and Late Flowering Are Two Terminal Flower1/Centroradialis Homologs That Control Two Distinct Phases of Flowering Initiation and Development in Pea. *Plant Cell*, vol. 15 (11), pp. 2742–2754 American Society of Plant Biologists. DOI: https://doi.org/10.1105/tpc.015701

Gallardo, K., Thompson, R. & Burstin, J. (2008). Reserve accumulation in legume seeds. *Comptes Rendus Biologies*, vol. 331 (10), pp. 755–762 No longer published by Elsevier. DOI: https://doi.org/10.1016/J.CRVI.2008.07.017

Gao, B., Bian, X.C., Yang, F., Chen, M.X., Das, D., Zhu, X.R., Jiang, Y., Zhang, J., Cao, Y.Y. & Wu, C.F. (2020). Comprehensive transcriptome analysis of faba bean in response to vernalization. *Planta*, vol. 251 (1), pp. 1–15 Springer. DOI: https://doi.org/10.1007/s00425-019-03308-x

Gao, M.J., Lydiate, D.J., Li, X., Lui, H., Gjetvaj, B., Hegedus, D.D. & Rozwadowski, K. (2009). Repression of Seed Maturation Genes by a Trihelix Transcriptional Repressor in Arabidopsis Seedlings. *The Plant Cell*, vol. 21 (1), p. 54 Oxford University Press. DOI: https://doi.org/10.1105/TPC.108.061309

Gazzarrini, S. & Song, L. (2024). LAFL Factors in Seed Development and Phase Transitions. *Annual review of plant biology*, vol. 75 (1), pp. 459–488. DOI: https://doi.org/10.1146/annurev-arplant-070623-111458

Gela, T.S., Khazaei, H., Podder, R., Vandenberg, A. & Gela, S. (2023). Dissection of genotype-by-environment interaction and simultaneous selection for grain yield and stability in faba bean (Vicia faba L.). *Agronomy Journal*, vol. 115, pp. 474–488. DOI: https://doi.org/10.1002/agj2.21268

Golombek, S., Rolletschek, H., Wobus, U. & Weber, H. (2001). Control of storage protein accumulation during legume seed development. *Journal of Plant Physiology*, vol. 158 (4), pp. 457–464 Urban & Fischer. DOI: https://doi.org/10.1078/0176-1617-00357

Grimberg, Å. (2019). Åkerbönan åter på tallriken. *Utsädesföreningens tidskrift*, vol. 01

Gulisano, A., Lippolis, A., van Loo, E.N., Paulo, M.J. & Trindade, L.M. (2023). A genome wide association study to dissect the genetic architecture of agronomic traits in Andean lupin (Lupinus mutabilis). *Frontiers in Plant Science*, vol. 13 (January), pp. 1–14. DOI: https://doi.org/10.3389/fpls.2022.1099293

Guo, B., Sun, L., Jiang, S., Ren, H., Sun, R., Wei, Z., Hong, H., Luan, X., Wang, J., Wang, X., Xu, D., Li, W., Guo, C. & Qiu, L.J. (2022). Soybean genetic resources contributing to sustainable protein production. *Theoretical and Applied Genetics 2022 135:11*, vol. 135 (11), pp. 4095–4121 Springer. DOI: https://doi.org/10.1007/S00122-022-04222-9

Gutiérrez, N., Pégard, M., Balko, C. & Torres, A.M. (2023). Genome-wide association analysis for drought tolerance and associated traits in faba bean (Vicia faba L.). *Frontiers in Plant Science*, vol. 14 (February), pp. 1–17

Gutierrez, N., Pégard, M., Solis, I., Sokolovic, D., Lloyd, D., Howarth, C. & Torres, A.M. (2024). Genome-wide association study for yield-related traits in faba bean (Vicia faba L.). *Frontiers in Plant Science*, vol. 15. DOI: https://doi.org/10.3389/fpls.2024.1328690

Gutierrez, N. & Torres, A.M. (2019). Characterization and diagnostic marker for TTG1 regulating tannin and anthocyanin biosynthesis in faba bean. *Scientific Reports*, vol. 9 (1) Nature Publishing Group. DOI: https://doi.org/10.1038/s41598-019-52575-x

Hayward, M.D. (1993). *Plant breeding : principles and prospects. Plant breeding : principles and prospects* 1. ed. London: Chapman & Hall. (Plant breeding series, 1)

Huber, J., Chaluppa, N., Voit, B., Steinkellner, S. & Killermann, B. (2023). Damage potential of the broad bean beetle (Bruchus rufimanus Boh.) on seed quality and yield of faba beans (Vicia faba L.). *Crop Protection*, vol. 168 (February),

p. 106227 Elsevier Ltd. DOI: https://doi.org/10.1016/j.cropro.2023.106227
Hufnagel, J., Reckling, M. & Ewert, F. (2020). Diverse approaches to crop diversification in agricultural research. A review. *Agronomy for Sustainable Development*. Springer. DOI: https://doi.org/10.1007/s13593-020-00617-4
Hughes, J., Khazaei, H. & Vandenberg, A. (2020a). Genetics of height and branching in faba bean (Vicia faba). *Agronomy*, vol. 10 (8). DOI: https://doi.org/10.3390/agronomy10081191
Hughes, J., Khazaei, H. & Vandenberg, A. (2020b). The Study of Genetics of Flower Color in Faba Bean Reveals Generous Diversity to Be Used in the Horticulture Industry. *HortScience*, vol. 55 (10), pp. 1584–1588 American Society for Horticultural Science. DOI: https://doi.org/10.21273/HORTSCI15238-20
Ijaz, U., Adhikari, K., Kimber, R., Trethowan, R., Bariana, H. & Bansal, U. (2021). Pathogenic Specialization in Uromyces viciae-fabae in Australia and Rust Resistance in Faba Bean. DOI: https://doi.org/10.1094/PDIS-06-20-1325-RE
Inci, N.E. & Toker, C. (2011). Screening and selection of faba beans (Vicia faba L.) for cold tolerance and comparison to wild relatives. *Genetic Resources and Crop Evolution*, vol. 58 (8), pp. 1169–1175 Springer. DOI: https://doi.org/10.1007/S10722-010-9649-2/TABLES/2
IPES-Food (2016). *From uniformity to diversity: a paradigm shift from industrial agriculture to diversified agroecological systems.* DOI: https://doi.org/10.1299/jsmedsd.2022.32.1208
Jayakodi, M., Golicz, A.A., Kreplak, J., Fechete, L.I., Angra, D., Bednář, P., Bornhofen, E., Zhang, H., Boussageon, R., Kaur, S., Cheung, K., Čížková, J., Gundlach, H., Hallab, A., Imbert, B., Keeble-Gagnère, G., Koblížková, A., Kobrlová, L., Krejčí, P., Mouritzen, T.W., Neumann, P., Nadzieja, M., Nielsen, L.K., Novák, P., Orabi, J., Padmarasu, S., Robertson-Shersby-Harvie, T., Robledillo, L.Á., Schiemann, A., Tanskanen, J., Törönen, P., Warsame, A.O., Wittenberg, A.H.J., Himmelbach, A., Aubert, G., Courty, P.E., Doležel, J., Holm, L.U., Janss, L.L., Khazaei, H., Macas, J., Mascher, M., Smýkal, P., Snowdon, R.J., Stein, N., Stoddard, F.L., Stougaard, J., Tayeh, N., Torres, A.M., Usadel, B., Schubert, I., O'Sullivan, D.M., Schulman, A.H. & Andersen, S.U. (2023). The giant diploid faba genome unlocks variation in a global protein crop. *Nature*, vol. 615 (7953), pp. 652–659. DOI: https://doi.org/10.1038/s41586-023-05791-5
Jensen, E.S., Peoples, M.B. & Hauggaard-Nielsen, H. (2010). Faba bean in cropping systems. *Field Crops Research*, vol. 115 (3), pp. 203–216. DOI: https://doi.org/10.1016/j.fcr.2009.10.008
Jezierny, D., Mosenthin, R., Sauer, N., Schwadorf, K. & Rosenfelder-Kuon, P. (2017). Methodological impact of starch determination on starch content and ileal digestibility of starch in grain legumes for growing pigs. *Journal of animal science and biotechnology*, vol. 8 (1) J Anim Sci Biotechnol. DOI: https://doi.org/10.1186/S40104-016-0131-7
Jia, H., Suzuki, M. & Mccarty, D.R. (2014). Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor

networks. *Wiley Interdisciplinary Reviews: Developmental Biology*, vol. 3 (1), pp. 135–145. DOI: https://doi.org/10.1002/wdev.126

Jo, L., Pelletier, J.M. & Harada, J.J. (2019). Central role of the LEAFY COTYLEDON1 transcription factor in seed development. *Journal of Integrative Plant Biology*, vol. 61 (5), pp. 564–580. DOI: https://doi.org/10.1111/jipb.12806

Kailash, C., Chand, S., Saini, R.P. & & Sharma, R. (eds.) (2017). *Smart Breeding : Molecular Interventions and Advancements for Crop Improvement,*. *The Charleston Advisor* Apple Academic Press, Incorporated. DOI: https://doi.org/10.5260/chara.19.2.39

Kang, Y., Li, M., Sinharoy, S. & Verdier, J. (2016). A snapshot of functional genetic studies in Medicago truncatula. *Frontiers in Plant Science*. Frontiers Media S.A. DOI: https://doi.org/10.3389/fpls.2016.01175

Karaköy, T., Toklu, F., Karagöl, E.T., Uncuer, D., Çilesiz, Y., Ali, A., Nadeem, M.A. & Özkan, H. (2023). Genome-wide association studies revealed DArTseq loci associated with agronomic traits in Turkish faba bean germplasm. *Genetic Resources and Crop Evolution*, pp. 1–18 Institute for Ionics. DOI: https://doi.org/10.1007/S10722-023-01615-7/TABLES/6

Karlgren, A., Gyllenstrand, N., Källman, T., Sundström, J.F., Moore, D., Lascoux, M. & Lagercrantz, U. (2011). Evolution of the PEBP Gene Family in Plants: Functional Diversification in Seed Plant Evolution. *Plant Physiology*, vol. 156 (4), pp. 1967–1977 Oxford Academic. DOI: https://doi.org/10.1104/PP.111.176206

Khamassi, K., Jeddi, F. Ben, Hobbs, D., Irigoyen, J., Stoddard, F., O'sullivan, D.M. & Jones, H. (2019). A baseline study of vicine-convicine levels in faba bean (Vicia faba L.) germplasm. DOI: https://doi.org/10.1017/S1479262113000105

Khan, M.A., Alghamdi, S.S., Ammar, M.H., Sun, Q., Teng, F., Migdadi, H.M. & Al-Faifi, S.A. (2019). Transcriptome profiling of faba bean (Vicia faba L.) drought-tolerant variety hassawi-2 under drought stress using RNA sequencing. *Electronic Journal of Biotechnology*, vol. 39, pp. 15–29 Pontificia Universidad Catolica de Valparaiso.

Khazaei, H., Carlson-Nilsson, U. & Schulman, A.H. (2024). The Jan Sjödin faba bean mutant collection: morphological and molecular characterization. *Hereditas*, vol. 161 (1), p. 37 BioMed Central. DOI: https://doi.org/10.1186/S41065-024-00339-7/FIGURES/4

Khazaei, H., O'Sullivan, D.M., Stoddard, F.L., Adhikari, K.N., Paull, J.G., Schulman, A.H., Andersen, S.U. & Vandenberg, A. (2021). Recent advances in faba bean genetic and genomic tools for crop improvement. *Legume Science*, p. leg3.75 John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/leg3.75

Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B.M., Watanabe, S., Yamada, T., Tabata, S., Kanazawa, A., Harada, K., Abe, J. & Kazusa, ; (2010). Two Coordinately Regulated Homologs of FLOWERING LOCUS T Are Involved in the Control

of Photoperiodic Flowering in Soybean 1[W][OA]. *American Society of Plant Biologists www.plantphysiol.org on August*, vol. 154, pp. 1220–1231. DOI: https://doi.org/10.1104/pp.110.160796

Krul, E.S. (2019). Calculation of Nitrogen-to-Protein Conversion Factors: A Review with a Focus on Soy Protein. *Journal of the American Oil Chemists' Society*, vol. 96 (4), pp. 339–364 John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/AOCS.12196

Landry, E.J. & Hu, J. (2019). Increasing pre-acclimation temperature reduces the freezing tolerance of winter-type faba bean (Vicia faba L.). *Journal of Agronomy and Crop Science*, vol. 205 (1), pp. 46–53 John Wiley & Sons, Ltd. DOI: https://doi.org/10.1111/JAC.12289

Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J. & Garnier, J. (2014). 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environmental Research Letters*, vol. 9 (10), p. 105011 IOP Publishing. DOI: https://doi.org/10.1088/1748-9326/9/10/105011

Lavania, D., Siddiqui, M.H., Al-Whaibi, M.H., Singh, A.K., Kumar, R. & Grover, A. (2015). Genetic approaches for breeding heat stress tolerance in faba bean (Vicia faba L.). *Acta Physiologiae Plantarum*, vol. 37 (1), p. 1737 Springer Berlin Heidelberg. DOI: https://doi.org/10.1007/s11738-014-1737-z

Leino, M.W. (2023). Vicia faba plant genetic resources preserved in home gardens in Sweden. *Legume perspectives*, (24), pp. 13–15

Link, W. (1990). Autofertility and rate of cross-fertilization: crucial characters for breeding synthetic varieties in faba beans (Vicia faba L.). *Theoretical and Applied Genetics*, vol. 79 (5), pp. 713–717 Springer-Verlag. DOI: https://doi.org/10.1007/BF00226888/METRICS

Link, W., Balko, C. & Stoddard, F.L. (2010). Winter hardiness in faba bean: Physiology and breeding. *Field Crops Research*, vol. 115 (3), pp. 287–296 Elsevier. DOI: https://doi.org/10.1016/J.FCR.2008.08.004

Lippolis, A., Hollebrands, B., Acierno, V., Jong, C. de, Pouvreau, L., Paulo, J., Gezan, S.A. & Trindade, L.M. (2025). GWAS Identifies SNP Markers and Candidate Genes for Off-Flavours and Protein Content in Faba Bean (Vicia faba L.). *Plants 2025, Vol. 14, Page 193*, vol. 14 (2), p. 193 Multidisciplinary Digital Publishing Institute. DOI: https://doi.org/10.3390/PLANTS14020193

Liu, W., Jiang, B., Ma, L., Zhang, S., Zhai, H., Xu, X., Hou, W., Xia, Z., Wu, C., Sun, S., Wu, T., Chen, L. & Han, T. (2018). Functional diversification of Flowering Locus T homologs in soybean: GmFT1a and GmFT2a/5a have opposite roles in controlling flowering and maturation. *New Phytologist*, vol. 217 (3), pp. 1335–1345 Blackwell Publishing Ltd.

Lu, X., Li, Q.T., Xiong, Q., Li, W., Bi, Y.D., Lai, Y.C., Liu, X.L., Man, W.Q., Zhang, W.K., Ma, B., Chen, S.Y. & Zhang, J.S. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *The Plant journal : for cell and molecular biology*, vol. 86 (6), pp. 530–544 Blackwell Publishing Ltd.

Lukanda, M.M., Dramadri, I.O., Adjei, E.A., Arusei, P., Gitonga, H.W., Wasswa, P., Edema, R., Ssemakula, M.O., Tukamuhabwa, P. & Tusiime, G. (2023). Genetic Diversity and Population Structure of Ugandan Soybean (Glycine max L.) Germplasm Based on DArTseq. *Plant Molecular Biology Reporter*, vol. 41 (3), pp. 417–426 Springer. DOI: https://doi.org/10.1007/S11105-023-01375-9/TABLES/4

Lyhagen, R. (2015). Plant breeding and variety development at the Svalöf and Weibull companies during 130 years (part 1). *Sveriges utsädesförenings tidskrift*

Lyhagen, R. (2016). Plant breeding and variety development at the Svalöf and Weibull companies during 130 years (part 3). *Sveriges utsädesförenings tidskrift*

Ma, W., Kong, Q., Arondel, V., Kilaru, A. & Bates, P.D. (2013). WRINKLED1, A Ubiquitous Regulator in Oil Accumulating Tissues from Arabidopsis Embryos to Oil Palm Mesocarp. *PLoS ONE*, vol. 8 (7), p. 68887. DOI: https://doi.org/10.1371/journal.pone.0068887

Maalouf, F., Abou-Khater, L., Babiker, Z., Jighly, A., Alsamman, A.M., Hu, J., Ma, Y., Rispail, N., Balech, R., Hamweih, A., Baum, M. & Kumar, S. (2022). Genetic Dissection of Heat Stress Tolerance in Faba Bean (Vicia faba L.) Using GWAS. *Plants*, vol. 11 (9), p. 1108 MDPI. DOI: https://doi.org/10.3390/PLANTS11091108/S1

Mackay, I. (2020). *Selection intensity. Optimizing breeding schemes. Manual.* Available at: excellenceinbreeding.org/toolbox/tools/eib-breeding-scheme-optimization-manuals%0AExcellenceinBreeding.org [2024-12-29]

MacNeill, G.J., Mehrpouyan, S., Minow, M.A.A., Patterson, J.A., Tetlow, I.J. & Emes, M.J. (2017). Starch as a source, starch as a sink: the bifunctional role of starch in carbon allocation. *Journal of Experimental Botany*, vol. 68 (16), pp. 4433–4453 Oxford Academic. DOI: https://doi.org/10.1093/JXB/ERX291

Mao, D., Michelmore, S., Paull, J., Preston, C., Sutton, T., Oldach, K., Yang, S.Y. & McMurray, L. (2019). Phenotypic and molecular characterisation of novel Vicia faba germplasm with tolerance to acetohydroxyacid synthase-inhibiting herbicides (AHAS) developed through mutagenesis techniques. *Pest management science*, vol. 75 (10), pp. 2698–2705 Pest Manag Sci. DOI: https://doi.org/10.1002/PS.5378

Martineau-Côté, D., Achouri, A., Karboune, S. & L'Hocine, L. (2022). Faba Bean: An Untapped Source of Quality Plant Proteins and Bioactives. *Nutrients*, vol. 14 (8), p. 1541 Multidisciplinary Digital Publishing Institute. DOI: https://doi.org/10.3390/nu14081541

Matias, I. (2024). *The Magical Fruit ? Ritual Aspects of Broad Beans in the Roman World In the Graduate College*. The university of Arizona.

Mawois, M., Vidal, A., Revoyron, E., Casagrande, M., Jeuffroy, M.H. & Le Bail, M. (2019). Transition to legume-based farming systems requires stable outlets, learning, and peer-networking. *Agronomy for Sustainable Development*, vol. 39 (1), pp. 1–14 Springer-Verlag France. DOI:

https://doi.org/10.1007/s13593-019-0559-1

Mbow, C., Rosenzweig, C., Barioni, L.G., Benton, T.G., Shukla, [P R, Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D.C., Zhai, P., Slade, R., Connors, S., Van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Pereira, J.P., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M. & Malley, J. (2019). Food Security. In: P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D.C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M., J.M. (ed.) *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*.

Mínguez, M.I. & Rubiales, D. (2021). Faba bean. *Crop Physiology Case Histories for Major Crops*, pp. 452–481 Academic Press.

Mossé, J. (1990). Nitrogen to Protein Conversion Factor for Ten Cereals and Six Legumes or Oilseeds. A Reappraisal of Its Definition and Determination. Variation According to Species and to Seed Protein Content. *Food Chem*, vol. 38, pp. 18–24. Available at: https://pubs.acs.org/sharingguidelines [2025-01-24]

Multari, S., Stewart, D. & Russell, W.R. (2015). Potential of Fava Bean as Future Protein Supply to Partially Replace Meat Intake in the Human Diet. *Comprehensive Reviews in Food Science and Food Safety*, vol. 14 (5), pp. 511–522 John Wiley & Sons, Ltd (10.1111). DOI: https://doi.org/10.1111/1541-4337.12146

Nikiforidis, C. V. (2019). Structure and functions of oleosomes (oil bodies). *Advances in Colloid and Interface Science*, vol. 274, p. 102039 Elsevier.

Nkhata, W., Shimelis, H., Melis, R., Chirwa, R., Mzengeza, T., Mathew, I. & Shayanowako, A. (2020). Population structure and genetic diversity analyses of common bean germplasm collections of East and Southern Africa using morphological traits and high-density SNP markers. *PLOS ONE*, vol. 15 (12), p. e0243238 Public Library of Science. DOI: https://doi.org/10.1371/JOURNAL.PONE.0243238

Nurmansyah, Alghamdi, S.S., Migdadi, H.M. & Farooq, M. (2019). Novel inflorescence architecture in gamma radiation-induced faba bean mutant populations. *International Journal of Radiation Biology*, pp. 1–8 Informa UK Limited.

De Pace, C., Delre, V., Scarascia Mugnozza, G.T., Maggini, F., Cremonini, R., Frediani, M. & Cionini, P.G. (1991). Legumin of Vicia faba major: accumulation in developing cotyledons, purification, mRNA characterization and chromosomal location of coding genes. *Theoretical and Applied Genetics*, vol. 83 (1), pp. 17–23 Springer-Verlag. DOI: https://doi.org/10.1007/BF00229221/METRICS

Panitz, R., Borisjuk, L., Manteuffel, R. & Wobus, U. (1995). Transient expression

of storage-protein genes during early embryogenesis ofVicia faba: synthesis and metabolization of vicilin and legumin in the embryo, suspensor and endosperm. *Planta 1995 196:4*, vol. 196 (4), pp. 765–774 Springer. DOI: https://doi.org/10.1007/BF01106772

De Pascale, S., Candido, V., Tuzel, Y., Savvas, D., Karkanis, A., Ntatsi, G., Lepse, L., Fernández, J.A., Vågen, I.M., Rewald, B., Alsin¸a, I., Alsin¸a, A., Kronberga, A., Balliu, A., Olle, M., Bodner, G., Dubova, L. & Rosa, E. (2018). Faba Bean Cultivation – Revealing Novel Managing Practices for More Sustainable and Competitive European Cropping Systems. *Frontiers in Plant Science*, vol. 9, p. 1115 Frontiers. DOI: https://doi.org/10.3389/fpls.2018.01115

Peoples, M.B., Hauggaard-Nielsen, H., Huguenin-Elie, O., Jensen, E.S., Justes, E. & Williams, M. (2019). The Contributions of Legumes to Reducing the Environmental Risk of Agricultural Production. *Agroecosystem Diversity*. Elsevier, pp. 123–143.

Quilichini, T.D., Gao, P., Yu, B., Bing, D., Datla, R., Fobert, P. & Xiang, D. (2022). The Seed Coat's Impact on Crop Performance in Pea (Pisum sativum L.). *Plants*, vol. 11 (15), p. 2056 MDPI. DOI: https://doi.org/10.3390/PLANTS11152056

Regeringen (2017). *En livsmedelsstrategi för Sverige − fler jobb och hållbar tillväxt i hela landet Regeringen.* Stockholm.

Röös, E., Carlsson, G., Ferawati, F., Hefni, M., Stephan, A., Tidåker, P. & Witthöft, C. (2020). Less meat, more legumes: Prospects and challenges in the transition toward sustainable diets in Sweden. *Renewable Agriculture and Food Systems*, vol. 35 (2), pp. 192–205

Sallam, A. & Ul-Allah, S. (2019). Genomics-Aided Breeding for Climate-Smart Traits in Faba Bean. *Genomic Designing of Climate-Smart Pulse Crops.* Springer International Publishing, pp. 359–395.

Segers, A., Dumoulin, L., Megido, R.C., Jacquet, N., Cartrysse, C., Kamba, P.M., Pierreux, J., Richel, A., Blecker, C. & Francis, F. (2022). Varietal and environmental effects on the production of faba bean (Vicia faba L.) seeds for the food industry by confrontation of agricultural and nutritional traits with resistance against Bruchus spp. (Coleoptera: Chrysomelidae, Bruchinae). *Agriculture, Ecosystems and Environment*, vol. 327 (December 2021)

Sharma, S., Upadhyaya, H.D., Varshney, R.K., Gowda, C.L.L., Jackson, S., Hayden, C.A. & Scaboo, A. (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. DOI: https://doi.org/10.3389/fpls.2013.00309

Sital, J.S., Malhotra, J.S., Sharma, S. & Singh, S. (2011). Comparative studies on biochemical components in mung bean [Vigna radiata (L.) Wilczek] varieties cultivated in summer and Kharif seasons. *Indian Journal of Agricultural Biochemistry*, vol. 24 (1), pp. 68–72

Sjödin, J. (1971). Induced morphological variation in Vicia faba L. *Hereditas*, vol. 67 (2), pp. 155–179 John Wiley & Sons, Ltd. DOI:

https://doi.org/10.1111/J.1601-5223.1971.TB02371.X

Skovbjerg, C.K., Angra, D., Robertson-Shersby-Harvie, T., Kreplak, J., Keeble-Gagnère, G., Kaur, S., Ecke, W., Windhorst, A., Kaergaard Nielsen, L., Schiemann, A., Knudsen, J., Gutierrez, N., Tagkouli, V., Ioana Fechete, L., Janss, L., Stougaard, J., Warsame, A., Alves, S., Khazaei, H., Link, W., Maria Torres, A., Martin, D.O., Uggerhøj Andersen, S., Rouf Mir, R., Kiel Skovbjerg cks, C. & Uggerhøj Andersen sua, S. (2023). Genetic analysis of global faba bean diversity, agronomic traits and selection signatures. *Theoretical and Applied Genetics 2023 136:5*, vol. 136 (5), pp. 1–27 Springer. DOI: https://doi.org/10.1007/S00122-023-04360-8

Skovbjerg, C.K., Knudsen, J.N., Füchtbauer, W., Stougaard, J., Stoddard, F.L., Janss, L. & Andersen, S.U. (2020). Evaluation of yield, yield stability, and yield–protein relationship in 17 commercial faba bean cultivars. *Legume Science*, vol. 2 (3), p. e39 John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/LEG3.39

Smits, M., Verhoeckx, K., Knulst, A., Welsing, P., de Jong, A., Houben, G. & Le, T.M. (2021). Ranking of 10 legumes according to the prevalence of sensitization as a parameter to characterize allergenic proteins. *Toxicology Reports*, vol. 8, pp. 767–773 Elsevier.

Song, Y., Wang, X.D. & Rose, R.J. (2017). Oil body biogenesis and biotechnology in legume seeds. *Plant Cell Reports 2017 36:10*, vol. 36 (10), pp. 1519–1532 Springer. DOI: https://doi.org/10.1007/S00299-017-2201-5

Stoddard, F. & Hämäläinen, K. (2011). Towards the world ' s earliest maturing faba beans. *Grain legumes*, (56), pp. 9–10

Stoddard, F.L., Nicholas, A.H., Rubiales, D., Thomas, J. & Villegas-Fernández, A.M. (2010). Integrated pest management in faba bean. *Field Crops Research*, vol. 115 (3), pp. 308–318 Elsevier.

Su, T., Wu, Y., Fang, C., Liu, B., Lu, S., Kong, F. & Liu, H. (2024). The Critical Roles of Phosphatidylethanolamine-Binding Proteins in Legumes. *Plant, Cell & Environment*,

Sukumaran, S., Rebetzke, G., Mackay, I., Bentley, A.R. & Reynolds, M.P. (2022). Pre-breeding Strategies. *Wheat Improvement: Food Security in a Changing Climate*, pp. 451–469 Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-90673-3_25/FIGURES/9

Szőke-Pázsi, K., Kruppa, K., Tulpová, Z., Kalapos, B., Türkösi, E., Gaál, E., Darkó, É., Said, M., Farkas, A., Kovács, P., Ivanizs, L., Doležel, J., Rabanus-Wallace, M.T., Molnár, I. & Szakács, É. (2024). DArTseq genotyping facilitates the transfer of "exotic" chromatin from a Secale cereale × S. strictum hybrid into wheat. *Frontiers in Plant Science*, vol. 15, p. 1407840 Frontiers Media SA. DOI: https://doi.org/10.3389/FPLS.2024.1407840/BIBTEX

Tanno, K.I. & Willcox, G. (2006). The origins of cultivation of Cicer arietinum L. and Vicia faba L.: Early finds from Tell el-Kerkh, north-west Syria, late 10th millennium B.P. *Vegetation History and Archaeobotany*, vol. 15 (3), pp. 197–204

Torkamaneh, D. & Belzile, F. (2022). *Genome-Wide Association Studies*. (Torkamaneh, D. & Belzile, F., eds.) 1. ed Quebec City, QC, Canada: Springer Nature. DOI: https://doi.org/https://doi.org/10.1007/978-1-0716-2237-7

UN (2017). *Goal 3: Ensure healthy lives and promote well-being for all at all ages — SDG Indicators*. Available at: https://unstats.un.org/sdgs/report/2017/goal-03/ [2020-08-12]

UN (2019). *Global Sustainable Development Report 2019: The Future is Now – Science for Achieving Sustainable Development*. New York.

UN (2020). *Goal 2: Zero Hunger – United Nations Sustainable Development*. Available at: https://www.un.org/sustainabledevelopment/hunger/ [2020-08-10]

UNCTAD (2024). *New UNCTAD-WHO analysis reveals trends in processed foods trade | UN Trade and Development (UNCTAD)*. Available at: https://unctad.org/news/new-unctad-who-analysis-reveals-trends-processed-foods-trade [2025-02-01]

Valamoti, S.M. (2023). *Plant Foods of Greece : A Culinary Journey to the Neolithic and Bronze Ages*. *Plant Foods of Greece : A Culinary Journey to the Neolithic and Bronze Ages* First edit. Tuscaloosa, Alabama: The University of Alabama Press. (Archaeology of Food Series)

Valdisser, P.A.M.R., Pereira, W.J., Almeida Filho, J.E., Müller, B.S.F., Coelho, G.R.C., de Menezes, I.P.P., Vianna, J.P.G., Zucchi, M.I., Lanna, A.C., Coelho, A.S.G., de Oliveira, J.P., Moraes, A. da C., Brondani, C. & Vianello, R.P. (2017). In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. *BMC Genomics*, vol. 18 (1), pp. 1–19 BioMed Central Ltd. DOI: https://doi.org/10.1186/S12864-017-3805-4/TABLES/1

Valente, I.M., Cabrita, A.R.J., Malushi, N., Oliveira, H.M., Papa, L., Rodrigues, J.A., Fonseca, A.J.M. & Maia, M.R.G. (2019). Unravelling the phytonutrients and antioxidant properties of European Vicia faba L. seeds. *Food Research International*, vol. 116, pp. 888–896 Elsevier Ltd.

Verdier, J., Kakar, K., Gallardo, K., Le Signor, C., Aubert, G., Schlereth, A., Town, C.D., Udvardi, M.K. & Thompson, R.D. (2008). Gene expression profiling of M. truncatula transcription factors identifies putative regulators of grain legume seed filling. *Plant Molecular Biology*, vol. 67 (6), pp. 567–580

Warsame, A.O., Michael, N., O'Sullivan, D.M. & Tosi, P. (2020). Identification and Quantification of Major Faba Bean Seed Proteins. *Journal of Agricultural and Food Chemistry*, vol. 68 (32), pp. 8535–8544

Warsame, A.O., Michael, N., O'Sullivan, D.M. & Tosi, P. (2022). Seed Development and Protein Accumulation Patterns in Faba Bean (Vicia faba, L.). *Journal of Agricultural and Food Chemistry*, vol. 70 (30), pp. 9295–9304

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., White, M., Khazaei, H., Ali, M., Street, D., Duc, G., Stoddard, F.L., Maalouf, F., Ogbonnaya, F.C., Link, W., Thomas, J. & O'Sullivan, D.M. (2016). A SNP-based consensus genetic map for synteny-based trait targeting in faba bean

(Vicia faba L.). *Plant Biotechnology Journal*, vol. 14 (1), pp. 177–185

Webb, A., Reynolds, T.R., Wright, T.I.C., Caiazzo, R., Lloyd, D.C., Thomas, J.E. & Wood, T.A. (2024). Identification of Faba bean genetic loci associated with quantitative resistance to the fungus Botrytis fabae, causal agent of chocolate spot. *Frontiers in Plant Science*, vol. 15 (April), pp. 1–14

Weber, H., Borisjuk, L. & Wobus, U. (2005). Molecular physiology of legume seed development. *Annual Review of Plant Biology*, vol. 56, pp. 253–279 Annual Reviews. DOI: https://doi.org/10.1146/annurev.arplant.56.032604.144201

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L.J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J.A., De Vries, W., Majele Sibanda, L., Afshin, A., Chaudhary, A., Herrero, M., Agustina, R., Branca, F., Lartey, A., Fan, S., Crona, B., Fox, E., Bignet, V., Troell, M., Lindahl, T., Singh, S., Cornell, S.E., Srinath Reddy, K., Narain, S., Nishtar, S. & Murray, C.J.L. (2019). Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, vol. 393 (10170), pp. 447–492 Elsevier.

Williams, O., Vander Schoor, J.K., Butler, J.B., Ridge, S., Sussmilch, F.C., Hecht, V.F.G. & Weller, J.L. (2022). The genetic architecture of flowering time changes in pea from wild to crop. *Journal of Experimental Botany*, vol. 73 (12), p. 3978 Oxford University Press. DOI: https://doi.org/10.1093/JXB/ERAC132

Yang, F., Chen, H., Liu, C., Li, L., Liu, L., Han, X., Wan, Z. & Sha, A. (2020). Transcriptome profile analysis of two Vicia faba cultivars with contrasting salinity tolerance during seed germination. *Scientific Reports 2020 10:1*, vol. 10 (1), pp. 1–10 Nature Publishing Group. DOI: https://doi.org/10.1038/s41598-020-64288-7

Yohane, E.N., Shimelis, H., Laing, M. & Shayanowako, A. (2022). Genetic diversity and grouping of pigeonpea [Cajanus cajan Millspaugh] Germplasm using SNP markers and agronomic traits. *PLOS ONE*, vol. 17 (11), p. e0275060 Public Library of Science. DOI: https://doi.org/10.1371/JOURNAL.PONE.0275060

Zanotto, S., Khazaei, H., Elessawy, F.M., Vandenberg, A. & Purves, R.W. (2020). Do Faba Bean Genotypes Carrying Different Zero-Tannin Genes (zt1 and zt2) Differ in Phenolic Profiles? *Journal of Agricultural and Food Chemistry*, vol. 68 (28), pp. 7530–7540 American Chemical Society. DOI: https://doi.org/10.1021/ACS.JAFC.9B07866/ASSET/IMAGES/LARGE/JF9B07866_0006.JPEG

Zhang, H., Fechete, L.I., Himmelbach, A., Poehlein, A., Lohwasser, U., Börner, A., Maalouf, F., Kumar, S., Khazaei, H., Stein, N. & Jayakodi, M. (2024). Optimization of Genotyping-by-Sequencing (GBS) for Germplasm Fingerprinting and Trait Mapping in Faba Bean. *Legume Science*, vol. 6 (3)

Zhang, H., Hu, Z., Yang, Y., Liu, X., Lv, H., Song, B.H., An, Y. qiang C., Li, Z. & Zhang, D. (2021). Transcriptome profiling reveals the spatial-temporal dynamics of gene expression essential for soybean seed development. *BMC*

Genomics, vol. 22 (1), pp. 1–13 BMC Genomics.

Zhao, J., Chen, J., Beillouin, D., Lambers, H., Yang, Y., Smith, P., Zeng, Z., Olesen, J.E. & Zang, H. (2022). Global systematic review with meta-analysis reveals yield advantage of legume-based rotations and its drivers. *Nature Communications 2022 13:1*, vol. 13 (1), pp. 1–9 Nature Publishing Group. DOI: https://doi.org/10.1038/s41467-022-32464-0

Zhao, N., Xue, D., Miao, Y., Wang, Y., Zhou, E., Zhou, Y., Yao, M., Gu, C., Wang, K., Li, B., Wei, L. & Wang, X. (2023). Construction of a high-density genetic map for faba bean (Vicia faba L.) and quantitative trait loci mapping of seed-related traits. *Frontiers in Plant Science*, vol. 14 (June)

Popular science summary

Faba bean (*Vicia faba* L.) is an ancient crop with remarkable potential for increased cultivation in modern agriculture—especially in Nordic regions with shorter growing seasons. Like other legumes, faba bean naturally improves soil fertility by capturing nitrogen from the air, reducing the need for artificial fertilizers. Its seeds are also exceptionally rich in protein, making it a valuable food source in the transition towards greater consumption of plant-based proteins. Increasing faba bean cultivation could help reduce reliance on imported soy, lower fertilizer use, and enhance biodiversity in agricultural systems.

Despite these benefits, faba bean has received considerably less breeding attention than crops like wheat and soybean, which have undergone extensive genetic improvements for yield and seed quality. To support the expansion of faba bean cultivation and its use as a food source, a better understanding is needed of the genetic mechanisms that regulate key traits such as productivity, seed composition, and climate adaptation. This knowledge can be applied to develop genomic tools that make plant breeding more efficient, contributing to a more resilient and sustainable food system while enhancing regional self-sufficiency.

This thesis characterized hundreds of faba bean varieties from around the world, evaluating their traits under Swedish field conditions to assess how well they adapt and perform in a Nordic climate. Through advanced genetic analysis, sections of the species' DNA were identified that influence critical traits such as plant height, flowering time, seed size, and seed yield. These genetic insights offer valuable guidance for future breeding efforts, enabling the development of faba bean varieties that are both more productive and better suited to northern growing conditions.

The research also explored, in detail, how faba bean seeds develop and store key nutrients such as protein, starch, and oil—essential components for both human food and animal feed. By understanding how these nutrients accumulate during seed development and identifying the genes that regulate their formation, improvements can be made to enhance the crop's nutritional value and its processing qualities for food production.

Another important finding concerns flowering, a key factor in the crop's adaptation to short growing seasons. The study identified a specific family of genes that regulate when and how faba bean plants flower.

By generating new genetic tools and insights for improving faba bean, this research supports the development of a more sustainable agricultural system with greater crop diversity and enhanced plant-based protein sources for the future.

Populärvetenskaplig sammanfattning

Åkerböna (*Vicia faba* L.) är en av världens äldsta odlade grödor och har stor potential att spela en viktig roll i framtidens jordbruk—särskilt i nordiska regioner där växtsäsongen är kort. Precis som andra baljväxter förbättrar åkerbönan jordens bördighet genom att binda kväve från luften, vilket minskar behovet av konstgödsel. Dess frön är dessutom rika på protein, vilket gör den till en värdefull livsmedelskälla i omställningen mot en mer växtbaserad kost. Om vi odlade mer åkerböna i Sverige skulle vi kunna minska vårt beroende av importerad soja och användningen av gödsel, samtidigt som den skulle öka mångfalden i våra odlingssystem. Trots sina många fördelar har åkerbönan fått relativt lite uppmärksamhet i växtförädlingen, särskilt i jämförelse med grödor som vete och soja, där omfattande forskning har lett till högre avkastning och bättre frökvalitet. För att kunna odla mer åkerböna och använda den i större utsträckning som livsmedel behöver vi bättre förstå hur dess gener styr viktiga egenskaper som avkastning, frökvalitet och anpassning till nordligt klimat. Genom att kartlägga de genetiska mekanismerna bakom dessa egenskaper kan vi utveckla nya verktyg som gör växtförädlingen mer effektiv och bidrar till ett mer hållbart livsmedelssystem, samtidigt som den regionala självförsörjningen stärks.

I den här avhandlingen har hundratals åkerbönesorter från hela världen undersökts för att utvärdera deras egenskaper under nordliga odlingsförhållanden. Med hjälp av avancerade genetiska analyser har delar av åkerbönans arvsmassa identifierats som påverkar viktiga egenskaper såsom växthöjd, blomningstid, fröstorlek och fröskörd. Dessa genetiska pusselbitar kan användas för att ta fram nya åkerbönesorter som är mer produktiva och bättre anpassade till nordliga odlingsförhållanden. Forskningen har också fokuserat på hur åkerbönans frön utvecklas och lagrar

viktiga näringsämnen som protein, stärkelse och olja—ämnen som är avgörande både för livsmedel och djurfoder. Genom att förstå hur dessa ämnen bildas under fröets tillväxt och vilka gener som styr processen kan vi förbättra åkerbönans näringsvärde och göra den mer användbar i livsmedelsproduktionen. En annan viktig del av forskningen handlar om blomningen, som är avgörande för hur väl åkerbönan klarar sig i vårt klimat. Vi har identifierat en särskild grupp av gener som bland annat påverkar när och hur åkerbönan blommar. Genom att ta fram ny kunskap och genetiska verktyg för att växtförädlingen av åkerböna bidrar den här avhandlingen till ett mer hållbart jordbruk, till större mångfald i odlingslandskapet och fler växtbaserade proteinkällor för framtiden.

Acknowledgements

To my main supervisor**, Per Hofvander**, and my almost-main supervisor**, Åsa Grimberg**— you've been my scientific compass through this journey, keeping me from wandering too far into the research wilderness. Your wisdom, gentle guidance, and remarkable ability to stay calm when I was convinced my experiments were conspiring against me have been invaluable.
To **Cecilia Hammenhag** and **Aakash Chawade**—though at a distance, your guidance has always arrived at just the right time, reminding me that help was always within reach.
To all my present and former colleagues at the **Plant Breeding department**, especially the **Biotech Group**—thank you. For the support, for the laughter, for the shared sighs of despair over fika, for the yoga sessions together. These years at SLU have been filled with science, yes, but also with warmth, camaraderie, and moments that made the lab feel less sterile.
To my **sister, Ellen**—your illustrations have brought this thesis to life, and I admire your depth and sharpness, and I am deeply grateful that you lent these skills to my thesis.
To those who made this work possible: **SLU Grogrund** for funding**,** and our collaboration partners **Lantmännen, Sveriges Stärkelseproducenter,** and **Kalmar Ölands Trädgårdsprodukter**—thank you for believing in this research, for supporting not just a project, but the pursuit of knowledge and progress. Special thanks to **Alf Ceplitis** for the many insightful discussions on quantitative genetics, which have not only deepened my understanding but also inspired me to engage with this field on yet another level.

This has been a journey of science, of learning, and of the people who made it meaningful. To all of you—thank you.

I

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 February 2024
DOI 10.3389/fpls.2024.1284997

OPEN ACCESS

EDITED BY
Chuanzhi Zhao,
Shandong Academy of Agricultural Sciences, China

REVIEWED BY
Hardy Rolletschek,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
Apoorv Tiwari,
National Agri-Food Biotechnology Institute, India

*CORRESPONDENCE
Hannah Ohm
hannah.ohm@slu.se
Åsa Grimberg
asa.grimberg@slu.se

†PRESENT ADDRESS
Ganapathi Varma Saripella,
CropTailor AB, Lantmännen, Svalöv, Sweden

RECEIVED 29 August 2023
ACCEPTED 18 January 2024
PUBLISHED 06 February 2024

CITATION
Ohm H, Saripella GV, Hofvander P and Grimberg Å (2024) Spatio-temporal transcriptome and storage compound profiles of developing faba bean (*Vicia faba*) seed tissues.
Front. Plant Sci. 15:1284997.
doi: 10.3389/fpls.2024.1284997

Spatio-temporal transcriptome and storage compound profiles of developing faba bean (*Vicia faba*) seed tissues

Hannah Ohm*, Ganapathi Varma Saripella†, Per Hofvander and Åsa Grimberg*

Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Lomma, Sweden

Faba bean (*Vicia faba*) is a legume grown in diverse climate zones with a high potential for increased cultivation and use in food due to its nutritional seeds. In this study, we characterized seed tissue development in faba bean to identify key developmental processes; from embryo expansion at the expense of the endosperm to the maturing storage stages of the bean seed. A spatio-temporal transcriptome profiling analysis, combined with chemical nutrient analysis of protein, starch, and lipid, of endosperm and embryo tissues at different developmental stages, revealed gene expression patterns, transcriptional networks, and biochemical pathways in faba bean. We identified key players in the LAFL (LEC1, ABI3, FUS3, and LEC2) transcription factor network as well as their major repressors VAL1 and ASIL1. Our results showed that proteins accumulated not only in the embryo but also in the endosperm. Starch accumulated throughout seed development and oil content increased during seed development but at very low levels. The patterns of differentially expressed transcripts encoding proteins with functions in the corresponding metabolic pathways for the synthesis of these storage compounds, to a high extent, aligned with these findings. However, the early expression of transcripts encoding WRI1 combined with the late expression of oil body proteins indicated a not manifested high potential for lipid biosynthesis and oil storage. Altogether, this study contributes to increased knowledge regarding seed developmental processes applicable to future breeding methods and seed quality improvement for faba bean.

KEYWORDS

***Vicia faba*, seed development, seed storage proteins (SSP), RNA-sequencing (RNA-seq), transcription factor (TF)**

1 Introduction

Vicia faba has several common names, including broad bean, horse bean, field bean, fava bean, and faba bean. It is an economically minor crop globally, with production quantities making out less than 2% of that of soybean which is the dominating grain legume today (FAOSTAT, 2021). Originally domesticated in North-West Syria around 10,000 BP, the faba bean is regarded as one of the oldest domesticated crops (O'Sullivan and Angra, 2016) with no identified wild ancestor. The main consumable is the fresh or dried seeds of faba bean that, due to their high protein content (25%-33% of dry-matter basis) (Heuzé et al., 2021) and balanced amino acid profile, are valuable as part of a healthy human diet as well as for animal feedstock (Ellwood et al., 2008; Multari et al., 2015). Increased cultivation of legumes such as faba bean, and their extended usage for human food consumption can support the ongoing dietary shift from animal to plant-based proteins (Röös et al., 2020). In addition, increased cultivation would contribute to a more sustainable food production system since legumes show low greenhouse gas- and water footprints (Semba et al., 2021) and have the ability to fertilize the soil by fixing atmospheric nitrogen, leading to a significant reduction of inputs in the agricultural sector (Lybæk and Hauggaard-Nielsen, 2019). Due to adaptation to diverse climate zones, faba bean also bears the potential to aid our transition to greater self-sufficiency of plant protein (Multari et al., 2015). Thus, it is a crop of nutritional and ecological value providing various ecosystem services (Punia et al., 2019), and can give an increased economic value if used to a higher extent in food products. Nevertheless, the biological processes behind nutrient storage and the regulatory network of gene expression during seed development of faba bean are still poorly investigated, the first paths being paved by Borisjuk et al. (1995; 2002). Indeed, knowledge about the genetic traits defining seed composition of stored macromolecules is fundamental for directed breeding and other approaches toward improved seed quality and quantity (Kang et al., 2016).

Generally, seed development can be split into the three phases histodifferentiation, seed filling, and desiccation. They span from the embryo's initial development and cell division to the exhaustion of the endosperm, and the accumulation of storage compounds in the cotyledons, until final seed desiccation and dormancy (Weber et al., 2005). The metabolic and regulatory pathways involved in these processes are well described for legume model species *Medicago truncatula* and *Lotus japonica*, as well as for other economically important legumes, such as soybean (*Glycine max*) and pea (*Pisum sativum*) (Gallardo et al., 2007; Dam et al., 2009; Malovichko et al., 2020; Sun et al., 2020). There is, however, a great variation in storage strategies within the legume family (Song et al., 2017), which motivates the need for research focusing on faba bean seed development. Due to the very large genome of the faba bean (13 Gb), with, until very recently, no reference genome available (Jayakodi et al., 2023), a transcriptomic approach of an RNA-sequencing (RNA-seq) analysis constituted an accessible method to study the genetic networks. This technique provides snapshots of gene expression in tissue and is a powerful tool for comparisons of differentially expressed transcripts (DETs), given certain circumstances such as treatments, varieties, or developmental stages.

The focus of this study was the major nutrient storage compartment of the faba bean seed, the embryo, and the tissue that surrounds it at an early developmental stage, the endosperm. For this we used two varieties that are commonly grown in Europe, one with white flower color (var. Taifun) and one with variegated flower color (var. Fanfare). Because of the absence of tannins in the seed coat, white flower colored faba bean types are often preferred for use in monogastric feed (Gutierrez et al., 2020), however yields are often lower than varieties with variegated flower color (Halling & Hagman, 2020). Our study did not place emphasis on pericarp tissue and variety variations in tannins because it is already known that genes encoding transcription factors (TFs) involved in the anthocyanin biosynthesis pathway control the presence of tannins in seed coat and flowers (Gutierrez et al., 2020).

Here, we combined transcriptome profiling with chemical analyses of storage compounds, to elucidate the major developmental stages for faba bean seed development and investigated the composition of key storage compounds in the embryo and endosperm tissues separately. This study elucidated the major metabolic shifts during seed filling of faba bean, with distinct phases of parallel processes occurring in the different tissues of the developing seed. To the best of our knowledge, it is the first of its kind to shed light on the transcriptional regulatory network underlying storage compound accumulation in this species.

2 Materials and methods

2.1 Plant material and growth conditions

Two commercially available faba bean varieties (var.), Taifun and Fanfare, with white and variegated flower colors respectively (Norddeutsche Pflanzenzucht, Germany), were cultivated in soil (50% peat, pH 5.5–6.5, added per m^3 soil: 5.5 kg lime, NPK 11-5-18 kg, 200 g micronutrients and 100 g iron) under greenhouse (for RNA-seq analysis) or Biotron (for metabolite analysis) conditions: 18°C –21°C light 6–22 h < 200 W/m^2, with a weekly supplement of Ca and NPK nutrient solution. Top shoots of plants were removed after a couple of weeks of growth to stimulate vegetative-reproductive competition and enhance flower retention for increased pod setting. Seeds were sampled at different developmental stages and different tissues (endosperm, embryo, pericarp, petals, and sepals) were isolated using a scalpel, spoon, and forceps. The sampled tissues were then snap-frozen in liquid nitrogen until stored at -80°C. Sampling was done continuously for 2–3 months after sowing until enough plant material of each developmental stage had been accessed, and sampling was done at approx. the same time point of the day to avoid any potential bias in gene expression due to circadian rhythm.

2.2 Tissue fixation and staining for microscopy

Longitudinal- and cross-sections of faba bean seeds at different developmental stages were fixed in 4% paraformaldehyde in Sorensens Phosphate Buffer (0.1M Na-phosphate, pH 7.0), shaken overnight, and rinsed in the same buffer the next day. Samples were stored in fresh phosphate buffer at 4°C until dehydrated in a series of EtOH concentrations: 30% for 3 h, 50% overnight, 70% for 3h, 95% for 3 h, and 100% EtOH for 24 h. Next, samples were cleared while shaking in a series of different EtOH-Xylene-solutions at different time periods: (2:1) overnight day one, (1:1) overnight day two, (0:1) two times for 3 h, and (0:1) for 24 h. Samples were embedded in melted paraffin (Histowax, Histolab, Gothenburg, Sweden) at 55°C for 3 x 24 h, cast into a mold, and stored dark at 4°C. Paraffin blocks were cut with Leitz Wetzlar 1515 rotary microtome (Leica Microsystems GmbH, Wetzlar, Germany) into 5–10 µm thin slices, and placed on fresh object glass (SuperFrost Plus, Menzel, Thermo Fisher Scientific, Waltham, Massachusetts, USA) at 37°C –40°C for 24–48 h, then stored at room temperature.

De-paraffinized samples were stained with Lugol's iodine (L) solution for starch detection, Light green (LG) for protein detection, and Sudan black B (SB) for lipid detection (Grimberg et al., 2020), and mounted with cover glass using Pertex mounting solution (Histolab, Gothenburg, Sweden). Microscopy images were acquired using a light microscope (Leica, DFC 450C camera, Wetzlar, Germany).

2.3 Determination of lipid, protein, and starch content

Seeds were sampled at various developmental stages, and embryo, endosperm, and pericarp were isolated, weighed, and freeze-dried for 48 h at 2.5 mbar and -60°C (Edwards Modulyo, UK). To obtain dry matter content, freeze-dried material was weighed before being stored at -80°C. Subsequent compound analysis was made on samples representing both varieties on four developmental stages (I-IV), with three biological replicates each. One biological replicate consisted of tissue pooled from 1–3 individual plants which were unique for that replicate.

For lipid extraction, 40 mg dry weight (dw) from each sample was extracted according to Bligh and Dyer (1959). An aliquot of 1 mL was concentrated in a glass tube on hot sand under nitrogen gas to a suitable volume and then applied on silica 60 thin layer chromatography plates of size 20 cm x 20 cm (Merck, Darmstadt, Germany) to separate lipid classes in heptane:diethyl ether:acetic acid (70:30:1 by vol.) by developing the plate in room temperature to its full length. After drying the plate and spraying it with primuline, free fatty acids (FFA) and triacylglycerol (TAG) species were located under UV light according to authentic standards. Identified regions of silica were scraped off and fatty acids in lipids were methylated into fatty acid methyl esters with 2% (v/v) sulphuric acid in dry methanol. Gas chromatography analysis of fatty acid methyl esters was performed on a CP-wax 58 column (FFAP-CB, 50 m, 0.32 mm inner diameter, 0.20 µm film, Varian, Palo Alto, USA) using an Agilent Technologies 8860 gas chromatograph (Santa Clara, US). Fatty acid methyl esters were quantified by using heptadecanoic acid methyl ester (Larodan, Solna, Sweden) as internal standard.

For protein and starch analyses, the freeze-dried material was first ground into fine flour in a mixer mill (MM 400, Retsch GmbH, Haan, Germany) at 30 Hz, using glass beads. For protein content determination, calculated from the total N residue, 3 mg flour was analyzed for total N quantification with the elemental analyzer (Flash 2000 Elemental Analyzer, Thermo Scientific). We followed the protocol for "Determination of total carbon and total nitrogen by dry combustion method (Dumas method) using a CN elemental analyzer" (Bieganowski et al., 2015), with Alfalfa as the reference standard (Krotz and Galotta, 2020). Subsequently, the 6.25 conversion factor converted total N content to protein content.

Total starch content was determined enzymatically on 40 mg flour with the Megazyme Total Starch HK Assay kit, K-TSHK (Megazyme, Brey, Ireland). Free sugars were first removed from flour by extraction with ethanol washes, followed by enzymatic degradation of starch according to the manufacturer's protocol, except for a down-scale of the colorimetric analysis step for the determination of glucose from degraded starch to fit a 96-well plate format. The absorbance was measured at 510 nm on a microplate spectrophotometer (Multiskan GO, ThermoFisher Scientific, MA, USA).

2.4 Sample preparation and RNA isolation

While frozen, plant tissues were ground in steel containers with 9–12 mm Ø steel beads chilled in N_2 to a fine powder using a mixer mill (MM 400, Retsch GmbH, Haan, Germany) at 30 Hz. RNA extraction was done following the protocol of RNeasy Mini Kit (Qiagen, Hilden, Germany), using the buffer RLC together with DTT. RNA extraction was done on endosperm and embryo tissue, representing both varieties in three developmental stages (I–III), with two or three biological replicates each (Supplementary Table 1). Additionally, RNA was extracted from pericarp tissue (developmental stage II), and petals and sepals for use in *de-novo* transcriptome assembly. RNA concentrations and RIN-values were measured with a NanoDrop (Thermo Fisher Scientific, Waltham, USA) and BioAnalyzer 2000 (Agilent Technologies, Santa Clara, USA). Samples were DNAse-treated with the TURBO DNA-free™ Kit (Invitrogen, MA, USA) before being sent for RNA sequencing.

2.5 RNA sequencing and quality control

Paired-end sequence reads were generated using Illumina high-throughput sequencing by Eurofins Genomics (Ebersberg, Germany), and initial Quality Control (QC) check was performed. Removal of ribosomal RNAs was done by aligning reads with SILVA (Quast et al., 2013) and Rfam (Kalvari et al., 2021) databases using Sortmerna-v2.1b (Kopylova et al., 2012), followed by TruSeq3 adapter trimming with Trimmomatic-v0.36 (Bolger et al., 2014) setting MINLEN:20 in bases and

SLIDINGWINDOW:5:20 along with other default parameters. The second round of QC checks was performed on independent samples with FastQC v0.11.7 (Andrews, 2010) and multiple sample visualization MultiQC v1.6 (Ewels et al., 2016).

2.6 *De-Novo* transcriptome assembly and functional annotation of differentially expressed transcripts

To perform *de-novo* transcriptome assembly, we used Trinity v3.2.2 workflow (Haas et al., 2013) and included all reads from all sequenced samples for the assembly. An additional SuperTranscripts script (https://github.com/trinityrnaseq/trinityrnaseq/wiki/SuperTranscripts) was applied to generate Trinity transcripts (Davidson et al., 2017). Transcript abundance was estimated using Salmon v1.3.0 (Patro et al., 2017). Raw read counts were used for Differential Expression analysis with DESeq2 (Anders and Huber, 2010; Anders et al., 2013) and in-built cross-sample "Relative Log Expression" (Love et al., 2014) normalization was performed by using minimum quality criteria. Further, two thresholds were set FDR < 0.05, and both FDR < 0.05 and Log2FoldChange (Log2FC) > 1.0, of which the most stringent was used for further downstream analysis. To annotate the function of expressed transcripts, the Trinity contig sequences were analyzed with blastx (NCBI-BLAST-v2.9.0+), parameters were set as -max_target_seqs 1, -evalue $1e^{-5}$ and searched against the Swiss-Prot/Uniprot database and scanned with TransDecoder-v5.5.0 (https://github.com/TransDecoder/TransDecoder) to predict open reading frames (ORFs). The predicted ORFs were further analyzed with the Trinotate-v3.2.2 pipeline (github.com/Trinotate/Trinotate.github.io). These ORFs were searched against Swiss-Prot/Uniprot database (uniprot_sprot) with blastp (-max_target_seqs 1, -e-value $1e^{-5}$) and then searched against other databases, such as, 1) Pfam-A.hmm release 34.0 using hmmscan v3.3, 2) TmHMM-2.0c, and 3) SignalP v.4.1. All the results were merged with Trinotate pipeline scripts, and an additional script (extract_GO_assignments_from_Trinotate_xls.pl) was used to generate Gene Ontology (GO) terms. Nucleotide sequences and corresponding ORFs of the generated transcripts were searched with blastx and blastp against predicted transcripts of three close relatives of faba bean, 1) *Medicago truncatula* (assembly: GCF_003473485.1), 2) *Pisum sativum* (Pea) (assembly: GCA_900700895.2) and 3) *Glycine max* (Soybean) (assembly: GCF_000004515.6). In addition plant model organism *Arabidopsis thaliana* (assembly: GCF_000001735.3) was included in comparative studies.

Plant-based TF databases such as PlantTFDB v5.0 (http://planttfdb.cbi.pku.edu.cn/) and iTAK v18.12 (Zheng et al., 2016) were included in the annotation. Database for plant model *Vicia faba* species is unavailable in both these TF databases, therefore other species such as *Medicago Truncatula*, *Glycine max*, and *Arabidopsis thaliana* databases were included for comparative analysis. Transcript expression levels were normalized with the Trimmed mean of M-values (TMM) method, due to the underlying distribution of expressed transcripts between samples being markedly different.

To highlight any sample variation we performed a principal component analysis (PCA) of the read count data. Quality filter was applied to transcripts and reduced to 41,821 by keeping transcripts that are expressed in at least two samples. Variance stabilizing transformation (VST) and plotPCA functions from the DESeq2 package where applied (Anders and Huber, 2010; Anders et al., 2013).

2.7 KEGG and GO enrichment analysis

The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed on the list of genes identified to be differentially expressed in all pairwise genotype comparisons, by using obtained gene coordinates from three model species *Arabidopsis thaliana*, *Medicago Truncatula*, and *Glycine max*. The KEGG Pathway enrichment analysis was performed with Gene Set Enrichment Analysis (GSEA) of KEGG, tested with submodule gseKEGG from the clusterProfiler (v3.18.1) R-package (Wu et al., 2021), with settings nPerm = 10,000, value of p cutoff = 0.01 and keeping remaining settings as default. The Gene Ontology (GO) over-representation test was performed using *Arabidopsis thaliana* coordinates, with enrichGO submodule from the clusterProfiler R-package (Yu et al., 2012) with settings pvalueCutoff = 0.05, qvalueCutoff = 0.05 and keeping other settings as default.

All scripts used in this study are available under: https://github.com/gvarmaslu/RNAseq_Faba-bean.

2.8 Statistical tests

Descriptive statistics were used to describe the characteristics of different seed tissue and developmental stages. Statistical analysis was performed on data of DETs and nutrient analyses by comparing means of development stages and tissues for each var. separately by a one-way analysis of variance (ANOVA) to test the significance of differences among samples followed by a posthoc Tukey's test at significance level $p \leq 0.05$ (Minitab 21.3.1, State College, PA, US). Means that do not share a letter in graphs are significantly different. Standard deviations (s.d) are indicated in graphs with error bars.

3 Results

3.1 Definition of seed tissue developmental stages

To identify different tissues in faba bean seeds, we performed a visual examination of beans that were dissected horizontally at three different developmental stages. Guided by previous descriptions of faba bean seed development (Patrick and Stoddard, 2010) we identified four distinct developmental stages, I to IV, used in this study (Figure 1A). These stages were determined based on visual assessment, according to the proportions of different seed tissues, i.e. of the endosperm and embryo. At stage I at ~15 days after

10.3389/fpls.2024.1284997

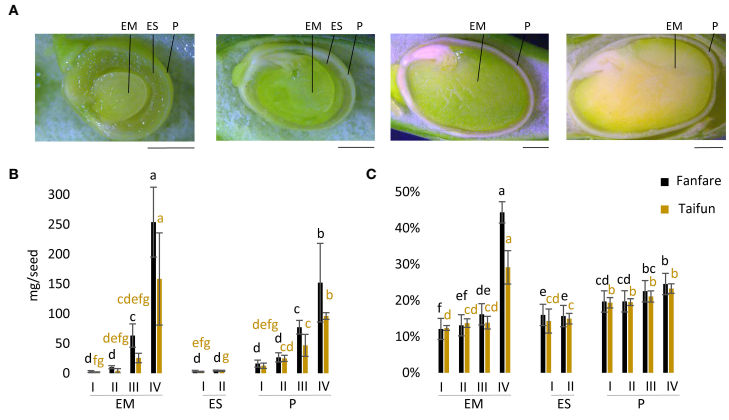


FIGURE 1
Microscopy images and dry matter content of faba bean seed developmental stages. **(A)** Microscopy images of longitudinal cut-open faba bean seeds at developmental stages I, II, III, and IV (from left to right). Scale bar: 2.5 mm. Dry weight of seeds in **(B)** mg/seed and **(C)** as % dry matter of embryo (EM), endosperm (ES), and pericarp (P) in var. Fanfare and Taifun at different developmental stages. Results are the mean ± s.d from three biological replicates. Means that do not share a letter are significantly different according to Tukey's test ($p < 0.05$), comparisons are made for each variety (color indicated) separately.

flowering (DaF), the earliest phase investigated here, the embryo was small but clearly visible, and surrounded by a relatively thick layer of endosperm. At stage II at ~17 DaF, the embryo was already occupying most of the seed space, surrounded by only a thin layer of endosperm. At stage III at ~32 DaF the seed had entered a desiccation phase and the color of the pericarp had changed from green to yellow-green, but with the embryo still being green. At this point, the bean was almost completely filled with cotyledons/embryo and the pericarp (seed coat) had stiffened. The final stage IV at ~45 DaF displayed a desiccating seed, and now also with the embryo/cotyledons shifting from green to yellow. In line with these visual observations, the size of the embryo and pericarp (mg dry matter per seed) increased during seed development, but not that of the endosperm (Figure 1B). There was a trend of larger embryo and pericarp size of the var. Fanfare as compared to var. Taifun at the later developmental stages (Figure 1B). Stage IV clearly comprised the desiccating, mature, and larger seed with significantly higher dw content of the embryo than its preceding immature stages (Figure 1C).

Histochemical staining and microscopy imaging of sections of fixed and paraffin-embedded seed parts served as additional anatomical references and as a confirmation of our definition of developmental stages. The outermost layers of the seed make out the seed coat, with the turquoise coloring of the light green (LG) staining, indicating protein-rich tissue and including the purple layer of the Lugol (L) staining, indicating a starch-rich section (Figures 2A–C). Going inward, the following layers of endosperm and embryo both showed to be protein-rich, with no or very little starch visible when double-stained in LG and L. Sudan black (SB) staining indicated low levels of lipids in the outer parts of the seed coat, and very low levels in the embryo (Figure 2D).

When comparing images of stained sections (Figure 2) with unstained sections (Supplementary Figures 1A–C), we noticed small dark-brown compartments present in the unstained sections of var. Fanfare that were not present in var. Taifun. Fanfare is a variety with variegated flowers that contain tannins in the seeds, which most likely explain the presence of these dark-brown substances since they could not be observed in sections from the white-flowering and tannin-free variety Taifun. The dark-brown compartments were present in two independent layers in the seed coat, and as part of cell compartments in the outer endosperm tissue of the faba bean, defining the border between (and including) the endosperm and seed coat. Tannic cell walls have been detected at the outer surface of the endosperm in *Arabidopsis thaliana* (Demonsais et al., 2020), and shown to concentrate in the central vacuole early postfertilization and subsequently oxidize, lending a brown hue to the seed coat in the seeds of *Arabidopsis thaliana* (Haughn and Chaudhury, 2005).

3.2 Characterization of seed starch, protein, and oil composition

To elucidate the nutrient storage strategy of the developing faba bean seed, storage compound levels were determined in the embryo at the four different developmental stages, and in the endosperm at the two earliest developmental stages (later stages had almost no visible endosperm tissue).

Protein levels were high and relatively stable and ranged from 31%–41% of the embryo dw, and were even higher, 41%–45%, of the endosperm dw, with a slightly decreasing trend in the embryo as the seed matured (Figure 3A). The opposite trend was shown for

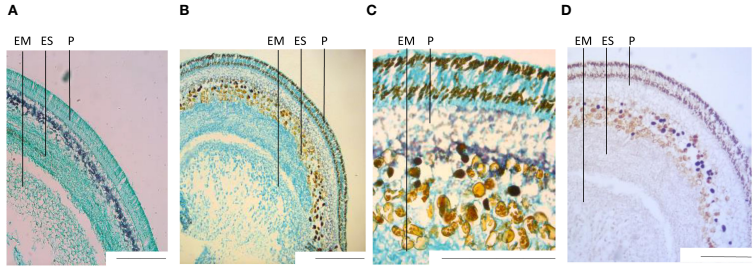


FIGURE 2
Transversal sections of fixed seeds, histochemically stained. Staining was done with Lugol's iodine solution (L) for starch detection, light green (LG) for protein recognition, and Sudan Black (SB) for lipid visualization. **(A)** var. Taifun and **(B)** var. Fanfare double stained with L+LG. **(C)** Close-up of Fanfare, double stained with L+LG. **(D)** Fanfare stained with SB. All images are from seed developmental stage I Scale bar: 2 mm. EM= embryo, ES= endosperm, P= pericarp.

both starch and TAG (oil) levels, which were increasing for every developmental stage with clear peaks of doubled values in stage IV as compared to stage III (Figures 3B, C). Starch levels ranged from as low as 2% up to 51% of embryo dw during seed development, but remained at a relatively low and stable level of 5%–9% of the endosperm dw. The level of TAG ranged between 0.11%–0.58% of the embryo dw and 0.11%–0.27% of the endosperm dw. The levels of FFA ranged between 0.02%–0.04% of the embryo dw and 0.01%–0.13% of the endosperm dw (Figure 3D). Except for protein levels, the endosperm was showing generally lower contents of the analyzed storage compounds than the embryo. The pattern of nutrient accumulation was similar for the two varieties studied.

3.3 Sequencing data analysis and functional annotation of transcripts

To explore the transcriptional dynamics of faba bean seed development, high-throughput next-generation transcriptome sequencing with Illumina technology (RNA-seq) was performed on three developmental stages and two different tissues, which on average resulted in 76 million paired-end raw reads per sample. After the adaptor sequences and low-quality sequences (about 10-12%) reads were removed, on average 68 million clean reads per sample were obtained. In total 2,350 million reads were used for *de-novo* transcriptome assembly (Supplementary Table 2). The Q30

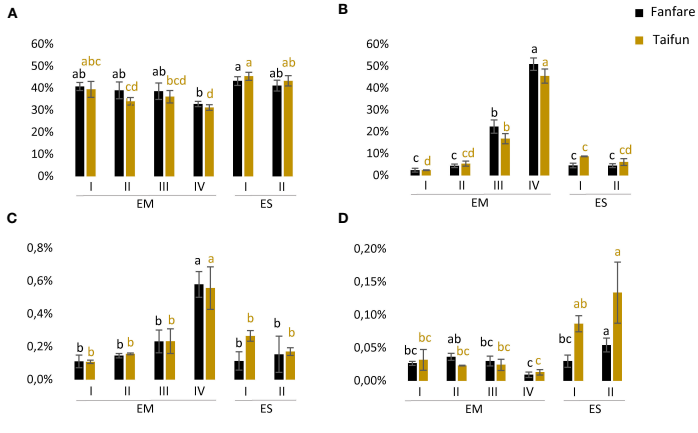


FIGURE 3
Accumulation of storage compounds in developing faba bean seeds. **(A)** Protein content (% by dw), **(B)** starch content (% by dw), **(C)** TAG content (% by dw), and **(D)** FFA content (% by dw) in faba bean embryo (EM) and endosperm (ES) at developmental stages I-IV. Results are the mean ± s.d from three biological replicates. Means that do not share a letter are significantly different according to Tukey's test ($p < 0.05$), comparisons are made for each variety (color indicated) separately.

percentage was 94% and the GC (Guanine and Cytosine) content was 41.7%. A total of 227,336 genes were predicted from *de-novo* assembly out of which 59,505 *de-novo* transcripts were predicted with open reading frames (ORFs). The transcripts were annotated by blast searches against local databases of three legume species *Glycine max* (soybean), *Medicago truncatula* (barrel clover), and *Cicer arietinum* (chickpea), as well as the plant model species *Arabidopsis thaliana* (thale cress), data is shown in Supplementary Table 3. This resulted in an average of 28,221 unique annotated transcripts with ORFs, out of which 14,821 were in agreement between all model species. A principal component analysis (PCA) of the read count data confirmed consistent grouping of biological replicates based on their tissue types, with the exception of one endosperm replicate (Supplementary Figure 2). Further, the PCA revealed distinct clustering among the three different developmental stages of the embryo. In contrast, the two developmental stages of the endosperm did not exhibit clear grouping. Notably, the varieties could not be clearly distinguished from each other in this analysis, which indicates high similarities between them.

3.4 Identification of differentially expressed transcripts during *Vicia faba* seed development

To get an overview of differences in gene expression between developmental stages of seed tissues, we determined the number of DETs in pairwise comparisons, based on the total 227,336 predicted genes (Figure 4; Supplementary Table 4; Supplementary Dataset 1). Generally, the biggest differences were found when comparing the latest with the earliest developmental stages, III vs I, in the embryo, for which around 6,000 transcripts were differentially expressed. A lower number of DETs were found between stage II vs I, than between stage III vs II. The number of DETs between endosperm tissue at stage II vs I was lower or equal to the number of DETs found in the embryo tissue at the same stages. DETs were also found between the white flowering var. Taifun and the variegated var. Fanfare, these differences were however not investigated further in this study. Instead, due to the high similarities found between the varieties in general, which are also seen in the abovementioned PCA (Supplementary Figure 2), the varieties are serving as biological replicates in this context.

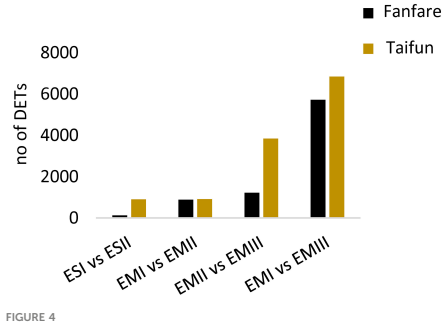


FIGURE 4
Differentially expressed transcripts in faba bean seeds. Comparison of the number of DETs between different developmental stages of seed tissue in faba bean var. Fanfare and Taifun. Further information on the pairwise comparisons at different thresholds is listed in Supplementary Dataset 1.

3.5 Differentially expressed transcripts encoding transcription factors in developing seed tissues

To understand the regulatory mechanisms of seed development, we investigated the list of transcripts that showed a differential expression during embryo and endosperm development and were annotated as plant TFs (Figure 5). They accounted for 915 out of the 10,217 DETs identified from all pairwise comparisons of embryo and endosperm tissue, as an average of the two var. Fanfare and Taifun. Of these TF families, those represented in the list of DETs to the highest extent were WRKY, NAC, bHLH, and C2H2, for both embryo and endosperm (Figures 5A, B; Supplementary Table 5). The embryo tissue showed a much higher number of DETs belonging to TFs (n=858), than the endosperm tissue (n=59). The expression levels of the most common DETs encoding TFs showed substantial variation, with the highest expression for the WRKY family in both endosperm and embryo tissue (Figure 5C).

To elucidate the connection between seed nutrient storage strategies and seed tissue development in faba bean, we looked further into transcripts annotated to specific TFs known to affect storage patterns in the embryo tissue. The LAFL network (LEAFY COTYLEDON1 (LEC1), ABSCISSIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3) and LEAFY COTYLEDON2 (LEC2)) is a set of TFs in plants that act as master regulators in seed development, triggering a cascade of secondary TFs, and regulating major hormone and signaling pathways (Lalanne et al., 2021). We found expressed transcripts annotated to encode three out of these four TFs in the developing faba bean embryo tissue (Figure 6A). Transcripts annotated to LEC1 showed decreasing expression with embryo development, while ABI3 was instead increasing. The expression level for the transcript annotated to FUS3 remained more or less constant throughout embryo development. Interestingly, transcripts homologous to LEC2 were not found in our dataset. Furthermore, transcripts annotated to other important TFs for embryo development or regulation of seed storage biosynthesis were looked at specifically (Figure 6B). The VIVIPAROUS1/ABI3-LIKE (VAL) TFs are known repressors of the LAFL network, enabling the transition from the embryonic to the vegetative state of the seedling (Jia et al., 2014). We found increasing expression of transcripts annotated as VAL1, but no DETs of VAL2, in the developing faba bean embryo. Transcripts annotated as the glycolysis and fatty acid synthesis regulating transcription factor WRINKLED1 (WRI1) were mainly expressed during the early stages of embryo development, with a significantly declining expression towards later stages. ASIL1 (for *Arabidopsis thaliana* 6b-interacting protein 1-like 1) contributes to the repression of the LAFL network by binding to its GT element (Gao et al., 2009). We found

10.3389/fpls.2024.1284997

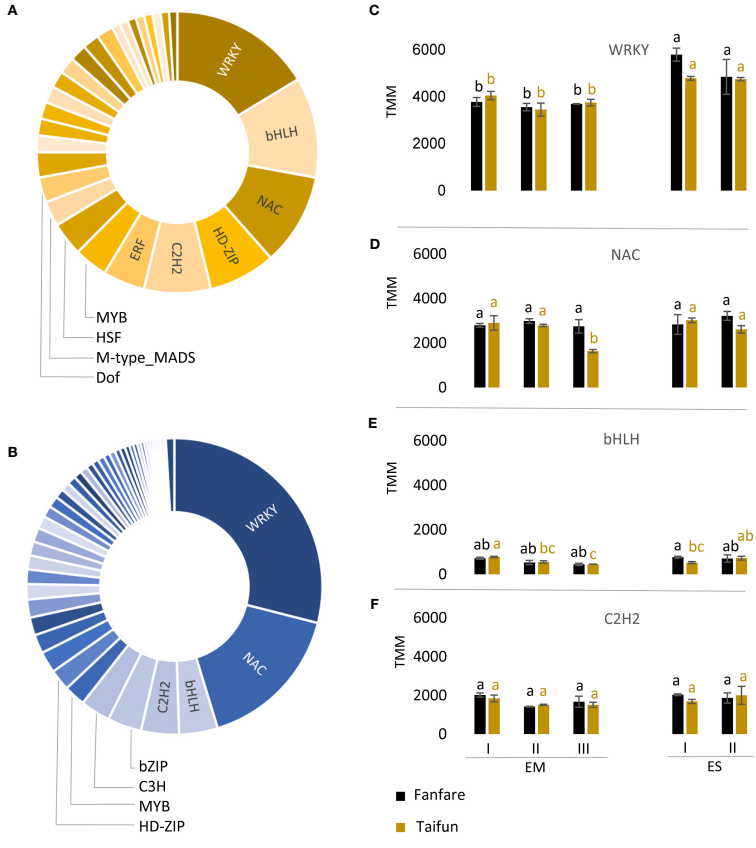


FIGURE 5
Transcription factor families shown to be differentially expressed in faba bean seeds during development. Pie charts of the distribution of DETs annotated to encode genes in TF families in **(A)** endosperm (yellow) and **(B)** embryo (blue) as an average between var. Fanfare and Taifun. **(C)** Expression levels for the DETs of the most common TF families in embryo (EM) and endosperm (ES) during the different developmental stages I to III were WRKY, **(D)** NAC, **(E)** bHLH, and **(F)** C2H2. Results are the mean ± s.d from three biological replicates. Means that do not share a letter are significantly different according to Tukey´s test ($p < 0.05$), comparisons are made for each variety (color indicated) separately.

transcripts annotated as ASIL1 to remain at stable expression levels in the two earlier stages of embryo development, but increasing towards developmental stage three. Transcripts of the ABI3-regulon of putative oleosins and late embryogenesis abundant proteins showed a clear increasing expression during embryo development.

3.6 Seed storage strategies in developing embryo

To get an overview of the transcriptional changes during embryo and endosperm development that are involved in central carbon metabolism, we identified DETs of selected genes encoding functions in storage protein, starch, and oil synthesis pathways (Figure 7A). KEGG enrichment scores for the relevant molecular pathways are depicted as a heatmap where blue and red colors indicate negative and positive differential expression over the developmental stages in each tissue (Figure 7B). In general, the expression patterns of the transcripts annotated as enzymes in the biosynthesis pathways of the major storage compounds were very similar for the two different varieties studied. Large changes were instead observed between different developmental stages in the embryo and only small changes were occurring in the endosperm.

Transcripts annotated as the seed storage proteins cupin, legumin, globulin, and vicilin were all highly differentially expressed in the seed tissues, with an increased expression during embryo development (Figure 7A). For starch and sucrose metabolism, the expression of transcripts was in general increasing in the embryo during seed development (Figure 7B). However, for the hexokinase that catalyzes the conversion of

10.3389/fpls.2024.1284997

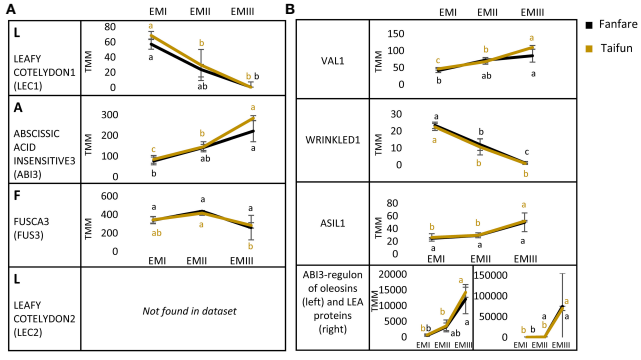


FIGURE 6
Differentially expressed transcripts (DETs) during faba bean embryo development, annotated to seed transcription factors (TFs). **(A)** Master regulator TFs, called the LAFL-network and **(B)** seed developmental transcription factor VAL1, fatty acid synthesis inducing transcription factor WRI1, ASIL1, putative ABI3-regulons part of the oleosin family and putative ABI3-regulons part of the late embryogenesis abundant (LEA) proteins. All expression levels are in TMM and show the var. Fanfare and Taifun. Results are the mean ± s.d of biological replicates. Means that do not share a letter are significantly different according to Tukey's test ($p < 0.05$), comparisons are made for each variety (color indicated) separately.

glucose to G6P, as well as fructose to F6P, the transcript expression trend was the opposite. For beta-amylase and glucose-1-phosphate adenyltransferase, the transcript expression trends in the embryo were different for the two varieties. Furthermore, while the transcripts annotated as beta-amylase and hexokinase were differentially expressed in the endosperm in Taifun, none of them showed differential expression in Fanfare. For the initial cytosolic steps of the glycolytic pathway, which regulate the conversion of sucrose to hexose phosphates, the faba bean embryo showed increased transcript expression during seed development. For the final steps of the plastidic glycolytic pathway, in which pyruvate is converted to acetyl–coenzyme A (CoA), which is a substrate for fatty acid synthesis in the plastid, the aggregated expression of transcripts was significantly lower than in the cytosol and was instead decreasing during embryo development. In the final steps of the starch and sucrose metabolism, leading towards the formation of starch, the faba bean embryo showed increasing transcript expression during embryo development.

In the glycerolipid metabolism in plants, fatty acids are first synthesized in the chloroplasts and later, after being carried into the endoplasmic reticulum as acyl-CoA, ester-linked with a hydroxyl group of glycerol to form TAGs that are stored in oil bodies (Schmid, 2021). All major enzymatic steps of the fatty acid biosynthesis pathway, except one, showed decreasing transcript expression in the faba bean embryo during seed development. A decreasing expression during embryo development was also observed for transcripts annotated as enzymes catalyzing a majority of the steps from acyl-CoA synthesis to TAG formation, and for transcripts of importance for the breakdown of TAGs to fatty acids and glycerol. The degradation of fatty acids in the peroxisomes, the beta-oxidation cycle, helps plants to recycle co-factors such as NADH or serves as a link to gluconeogenesis. For transcripts important for the beta-oxidation cycle, we observed inconsistent results for the two varieties but with a tendency towards the enrichment of transcripts in the later stages of embryo development. Surprisingly, the expression of transcripts annotated as encoding plant oil body proteins, oleosins, was on an opposite trajectory (as compared to the observed decreasing trend for fatty acid and TAG synthesis), being highly expressed in the later developmental stages of the embryo.

4 Discussion

4.1 More complex developmental switches in the embryo than in the endosperm

The seed development of faba bean is coordinated by interactions of the seed coat, endosperm, and embryo tissues, and is mainly driven by genetic programs, environmental factors, hormones, and the transport and uptake of photoassimilate sugars. Our distinction into four developmental stages based on visual assessment of horizontally dissected seeds, used throughout this paper, indicated major changes throughout the seeds' development and maturation with regards to the increased proportion of embryo and the concomitantly decreased proportion of endosperm, as well as the increasing dry matter content. This could be confirmed by significant differences during seed development in both nutrient levels and levels of DETs annotated to encode many important metabolic and developmental markers. In the early seed developmental stages of faba bean, nutrients are stored in the endosperm and from there distributed to the developing embryo (Patrick and Stoddard, 2010). The endosperm tissue is therefore clearly visible at this stage, forming a jelly-like layer between the embryo and seed coat. The faba bean is an exalbuminous seed, and after the replenishment of the endosperm, its function as transient nutrient storage is replaced by the inner seed coat (Patrick and Stoddard, 2010).

10.3389/fpls.2024.1284997

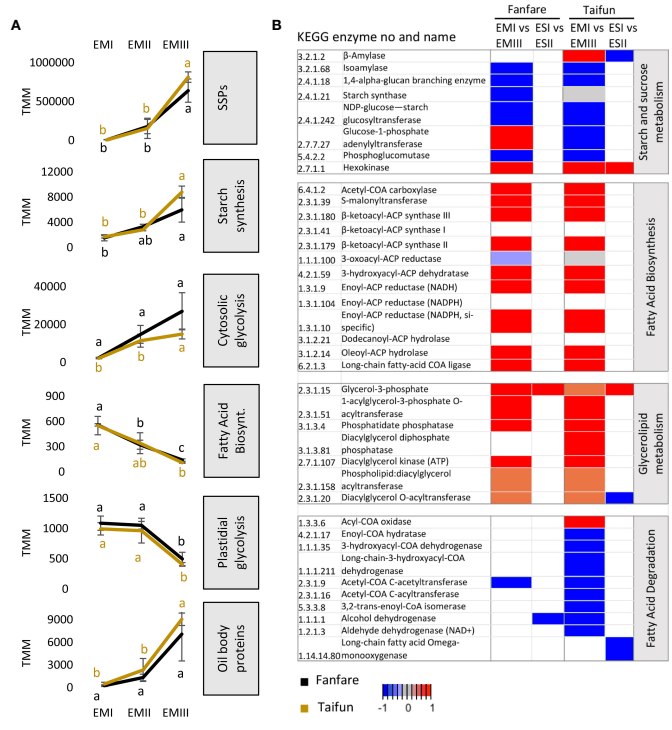


FIGURE 7
Differential expression of transcripts encoding proteins involved in major metabolic pathways of storage compounds. **(A)** Aggregated expression levels (TMM) of selected groups of differentially expressed transcripts during embryo development in the var. Fanfare and Taifun. Gene names and transcript IDs for the homologs used in the graphs are listed in Supplementary Dataset 2. **(B)** Heatmap of selected genes from enriched KEGG pathways of four major metabolic pathways for the biosynthesis of storage compounds. The expression pattern is shown as a relative color scheme for each row, comparing the different developmental stages of embryo and endosperm tissues in faba bean varieties Taifun and Fanfare. Red is indicating decreasing expression (higher expression at early compared to late developmental stage) and blue is indicating increasing expression (lower expression at early compared to late developmental stage). Results are the mean ± s.d of biological replicates. Means that do not share a letter are significantly different according to Tukey's test ($p < 0.05$), comparisons are made for each variety (color indicated) separately.

During this process, the embryo grows fast and finally takes up most of the seed space. The early developmental stages I and II in our study had separable embryo and endosperm tissues and were both phases of nutrient synthesis and accumulation. Stage III and IV instead showed a depleted endosperm and were phases of seed filling, nutrient storage, and desiccation. The larger embryo size in var. Fanfare as compared to Taifun during later developmental stages aligns with the available data on thousand grain weight of mature seeds obtained from variety testing in field (Halling & Hagman, 2020).

The embryo tissue showed a higher number of DETs during seed development belonging to TFs than the endosperm tissue, which is indicative of more complex developmental and metabolic switches in the embryo as compared to the endosperm. This in turn, was in line with the different patterns of storage compound accumulation during seed development observed between tissues in our study, with the levels of starch and oil increasing in the embryo, but in principle staying at a constant level in the endosperm. It was interesting to note that the levels of protein, as determined from total nitrogen, were kept at high and relatively constant levels in both tissues during the first three stages of seed development. The two var. Taifun and Fanfare showed similar spatiotemporal patterns in embryo and endosperm of both storage compound accumulation and of DETs involved in seed development and storage compound biosynthesis, and the varieties could therefore serve as confirmative biological replicates. The variety difference in tannin content highlighted a possible anatomical border between the endosperm and the seed coat based on our light microscopy analysis of fixed seed tissues, with clearly visible dark-brown compartments in the outer cell layers of non-stained slices of seeds of Fanfare, which were absent in seeds of the white-flowering var. Taifun. Although not previously described in faba bean, studies in *Arabidopsis thaliana* are indicative of similar structures containing tannins (Haughn and Chaudhury, 2005; Demonsais et al., 2020).

4.2 LAFL-network is regulating early embryo development

Nearly 7% of the genes in higher plants encode TFs, which are proteins with a specific DNA-binding domain that is regulating transcription (Le et al., 2011). Many fundamental parts of plant development are regulated through the activation and repression of target genes by TFs (Verdier et al., 2008). From the analysis of RNAseq data, we could note a similar proportion of TFs (6%) in the list of differently expressed transcripts during the faba bean seed development. Among those, we identified several transcripts annotated as encoding TFs that have been characterized as important in *Arabidopsis thaliana* seed development. The most common occurring transcripts among the identified DETs annotated as TFs in faba bean endosperm and embryo development, all belonged to the biggest TF families in plants and are associated with seed development in several other plant species. For example, we found that WRKY was the most abundant TF-family differentially expressed in both endosperm and embryo tissue during the faba bean seed development. From previous studies on several different plant species, including closely related *Medicago truncatula*, the LAFL network is found to be an important set of TFs that act as master regulators in seed development and the deposition of storage reserves such as starch, lipids, and SSPs (Kagaya et al., 2005; Lalanne et al., 2021). However, very little is known about the LAFL network in faba bean. Our results showed that transcripts homologous to three out of the four TFs part of the network were present in the developing embryo tissue of the faba bean. LEC1 showed decreasing expression during embryo development, ABI3 was increasing and FUS3 remained constant. Interestingly, transcripts homologous to LEC2 were neither found in our dataset nor in the predicted open reading frames of the *Vicia faba* reference genome Hedin/2 (Jayakodi et al., 2023). The trends seen in our data follow prior suggestions that LEC1 is a pioneering TF and primarily functions during early seed development stages (embryogenesis), whereas FUS3 and ABI3 activate maturation-specific processes and thus are mostly involved in later stages of seed development (Schneider et al., 2016). The TFs VAL1 and ASIL1, in turn, act as a repressor of the LAFL network (Jia et al., 2014). This is in accordance with our data where the transcripts homologous to VAL1 and ASIL1 were expressed in an opposite pattern to that of LEC1 in the faba bean embryo during seed development. Future studies of individual genes within these sets are of interest to validate our findings and gain a better understanding of their specific roles in seed development, for example using techniques such as quantitative real-time PCR.

4.3 Oil bodies as a putative temporal energy reserve in embryonic tissue

Faba bean is distinguished by a combination of its relatively high protein content (30%) and low-fat content (3%) in mature seeds, in comparison with many other grain legumes (Song et al., 2017). In the chloroplasts and other plastids of the plant cell, fatty acids are *de novo* synthesized from acetyl–CoA to be further distributed to other cell compartments where they can be used for the synthesis of different glycerolipids. Glycerolipids can act as structural components of the cell (membrane lipids) and can serve as energy storage in the form of TAGs (Baud and Lepiniec, 2010). FFAs and glycerol can be converted into TAG as part of the process of lipid metabolism and conversely, when energy is needed, TAG can be digested into FFA and glycerol through lipolysis. Our nutrient analysis revealed that, even though being at a low concentration as compared to other major storage compounds, the level of TAGs (oil) was increasing substantially throughout the seeds' development, with doubled concentration from stage I to stage III in embryo tissue and a five times higher concentration in stage IV. FFA levels, on the other hand, were low and decreasing during embryo development, possibly for use in TAG biosynthesis or as part of the structural formation of the cells' membrane formations. The transcription factor WRI1 is known to induce oil synthesis during seed embryo maturation by activating the transcription of genes in glycolysis and *de novo* fatty acid synthesis in *Arabidopsis thaliana*, as well as in several other species, including soybean (Focks and Benning, 1998; Ma et al., 2013; Chen et al., 2018). At an early embryo developmental stage, our transcriptomic data of faba bean showed a high expression of WRI1, as well as of its known target genes in fatty acid synthesis and the earlier parts of plastidic glycolysis (which feeds fatty acid synthesis with carbon precursors), indicating a positive regulation of metabolism for the production of FFAs to be used for TAGs (Grimberg et al., 2015). The same trend of high expression levels early and lower expression levels later during seed development was seen for transcripts encoding enzymes in glycerolipid metabolism that leads to TAG synthesis. Oleosomes, also called lipid droplets or oil bodies, are organelles for the storage of TAG. A matrix of accumulated TAGs is surrounded by a protective phospholipid monolayer, including oil body proteins (Miray et al., 2021). The transcripts encoding oil body proteins showed increased levels as the faba bean embryo developed, which is indicative of increasing oil accumulation. Accordingly, the transcripts for ABI3, a TF known in other plant species to regulate the expression of genes encoding oleosins (Lalanne et al., 2021), was also showing increasing levels during embryo development. The observed increase of TAG levels as well as transcripts encoding for oil body proteins and ABI3, in combination with the opposite pattern of the transcripts encoding for WRI1 and fatty acid biosynthesis, show similarities to the Arabidopsis seed (Ruuska et al., 2002). A decreasing trend of *WRI1* expression during seed development has also been reported in several oil-seed species (Troncoso-Ponce et al., 2011). Intriguingly, despite these similarities to faba bean, the levels of oil in mature seeds of Arabidopsis are much higher. However, the onset of lipid droplet organizing proteins, such as oleosins, have a much higher degree of gene expression overlap with *WRI1* in oil seeds during development. It could therefore be possible that lipids, in the absence of organization as TAG oil bodies by lipid droplet proteins, act as temporal energy storage in faba bean seeds and are turned over during development. However, further biochemical characterization is needed to support such hypothesis.

4.4 Proteins are accumulated not only in the embryo but also in the endosperm tissue

Although structurally different among different plant species, seed storage proteins (SSPs) share the characteristics of accumulating to high levels in specific seed-organ tissues at certain time points of the seed development (Krishnan and Coe, 2001) and are mainly regulated at the transcriptional level (Verdier et al., 2008). Our study showed that the total protein content, based on total nitrogen determination, was high in the embryo tissue already at an early stage of development (40% by dry weight) and then decreased slightly towards seed maturation (31-33% by dry weight) and that the level was even slightly higher in endosperm tissue. The seed protein content in both varieties were 28-29% as determined from variety tests in field (Halling and Hagman, 2020), which is just below the content in embryo tissue at our latest developmental stage analyzed. A higher seed protein content in the early developing stages as compared to later stages has previously been observed in other investigations of faba bean seeds (Warsame et al., 2022) as well as in other legume seeds (Sital et al., 2011; Zhang et al., 2021). However, it should be noted that total protein levels, estimated through total nitrogen determinations, do not exclusively represent storage proteins. This is because a portion of the total nitrogen comes from structural proteins, enzymes, amides, free amino acids etc. (Ezeagu et al., 2002). This nitrogen portion can be expected to be more prominent during the early stages of seed development, characterized by rapid cell division, as compared to later stages when there is a substantial accumulation of storage proteins. Nevertheless, light green histochemical staining, which stains proteins by binding to free basic side chains of a protein by its sulfonic acid group (Oud et al., 1984), of fixed and sliced seed tissues confirmed the presence of high protein levels in the early phases of both embryo and endosperm development.

The main storage proteins in faba bean are globulins, which are comprised of legumin and vicilin/convicilin, accounting for almost half and one-third of the SSPs respectively (Warsame et al., 2020). They are known to be synthesized at 21-28 DaF in faba bean (Patrick and Stoddard, 2010), time points that within this study match phases II and III of seed development. The transcriptomic data showed an increasing expression for transcripts annotated as SSPs during embryo development, which is in line with an increasing protein content in absolute terms (mg protein/seed).

4.5 Starch is accumulating in the later stages of embryo maturation

Starch makes out the major part of the carbohydrates found in the mature faba bean seed, with levels between 30%–42% of dw (Cerning and Guilbot, 1975). During the day, photosynthates are exported from the chloroplast and partitioned as sucrose in the phloem, either for immediate uses (growth assimilates), temporary storage as transient starch in the cytosol, or imported to amyloplasts for long-term energy storage, so-called storage starch (MacNeill et al., 2017). With the help of histochemical staining, we could visually show that the seed coat at early seed developmental stages is rich in amyloplasts, organelles that synthesize and store starch, later on providing a possible carbon source for the maturing embryo (Weber et al., 2005). Our nutrient analysis showed that the embryo, in turn, accumulates starch increasingly towards the later developmental stages which could further be confirmed by our transcriptome results showing an increasing trend of transcripts levels encoding enzymes in the cytosolic glycolytic pathway. As seed size stabilizes and the endosperm is depleted, the cytosolic glycolytic activity increases to support starch and protein storage. In glycolysis, plants oxidize hexoses to generate ATP and organic acids and produce building blocks for anabolism. This is in agreement with our nutrient analysis which showed the highest levels of starch content in the later stages (III and IV) of seed development, as well as with our increasing transcriptomic data of homologous genes annotated to encode starch synthase, 1,4-alpha-glucan branching enzyme, and ADP-glucose pyrophosphorylase, all playing important roles in the starch biosynthesis pathway in plants (Stitt and Zeeman, 2012).

4.6 Conclusion and future directions

Seed quality is an important breeding target in faba bean to nourish an increasing population with better and more sustainably produced food and animal feed. The complex relationship between seed development, nutrient biosynthesis, and the resulting seed quality of faba bean is therefore an important area of study that can be of interest to the whole value chain of faba bean, from plant breeding to food and feed producers and consumers. Increased knowledge of the genetic regulation of synthesis and storage patterns of protein, starch, and oil of seeds is an important step towards identifying specific markers that can be used in efficient breeding of protein-rich legumes. In this study, we identified highly differentially expressed transcripts during seed development in faba bean that were annotated as encoding enzymes in the synthetic pathways for these storage compounds, as well as TFs known to regulate embryo development. This can form the basis for further research to identify breeding targets for desired seed qualities in faba bean. One interesting suggested breeding target is the subunit composition of storage proteins towards a higher legumin:vicilin ratio which could enhance the faba bean's nutritional qualities significantly with regards to amino acid composition (Warsame et al., 2020). A somewhat surprising finding of the current study of faba bean was that the expression pattern of genes involved in fatty acid and oil biosynthesis was similar to that in high-oil accumulating plant species. Therefore, another interesting breeding target could be to explore possibilities of increasing the oil content from the current very low level in faba bean, having in mind that soybean is today a dual-purpose legume crop with a high economic value for both oil and protein (Guo et al., 2022). In starch-rich grain legume crops, such as faba bean, there is usually a trade-off towards oil. Varieties with lower starch content could therefore show a higher capacity to store

lipids, a useful trait for further biotechnological modifications toward increased oil content (Song et al., 2017).

Data availability statement

The datasets for this study can be found in the Sequence Read Archive (SRA) database at NCBI under BioProject accession number PRJNA861904. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA861904.

Author contributions

HO: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. GS: Data curation, Formal analysis, Methodology, Software, Writing – review & editing. PH: Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing – review & editing. ÅG: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was made possible by: SLU Grogrund Center for Breeding of Food Crops; Future faba beans for food and feed (#ltv.2018.1.1.1-1067). PlantLink and SLU Bioinformatics Infrastructure for their bioinformatics support. Martha & Dagny Larssons fond (#212-214) for financing RNA-seq. The Royal Physiographic Society in Lund, Sweden, for financial support for analytical equipment. Uppsala University at UPPMAX for providing the projects SNIC 2021/23-648 and SNIC 2021/5-533 for data handling. SLU LTV faculty for supporting the growth facilities.

Acknowledgments

We thank Jenny Östberg for assistance in sampling the seeds, Anna Nieto-Esteve for assistance in starch analysis and Salla Marttila for advice in microscopy analysis. We also thank the greenhouse/Biotron personnel at LTV faculty for assistance during the experiments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1284997/full#supplementary-material

SUPPLEMENTARY MATERIAL
Supplementary Figures 1–2 and Supplementary Table 1–5.

SUPPLEMENTARY DATASHEET 1
List of all differentially expressed transcripts from all pairwise genotype comparisons.

SUPPLEMENTARY DATASHEET 2
AT Locus names and search queries for transcripts used in Figure 7A.

References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, 1–12. doi: 10.1186/GB-2010-11-10-R106/COMMENTS

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., et al. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* 8, 1765–1786. doi: 10.1038/nprot.2013.099

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data [Online]. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Baud, S., and Lepiniec, L. (2010). Physiological and developmental regulation of seed oil production. *Prog. Lipid Res.* 49, 235–249. doi: 10.1016/J.PLIPRES.2010.01.001

Bieganowski, A., Stanislav, M, Frąc, M., Tuf, I. H., Brzezińska, M. V. M., Siebielec, G., et al. (2015). *Soil analysis Laboratory manual.* 1–119. Available at: http://projekty.ipan.lublin.pl/uploads/laboratory_manual.pdf.

Bligh, E. G., and Dyer, W. J. (1959). A Rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37, 911–917.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Borisjuk, L., Walenta, S., Rolletschek, H., Mueller-Klieser, W., Wobus, U., and Weber, H. (2002). Spatial analysis of plant metabolism: Sucrose imaging within *Vicia faba* cotyledons reveals specific developmental patterns. *Plant J.* 29, 521–530. doi: 10.1046/j.1365-313x.2002.01222.x

Borisjuk, L., Weber, H., Panitz, R., Manteuffel, R., and Wobus, U. (1995). Embryogenesis of *vicia faba* L.: histodifferentiation in relation to starch and storage protein synthesis. *J. Plant Physiol.* 147, 203–218. doi: 10.1016/S0176-1617(11)81507-5

Cerning, S., and Guilbot, (1975). Carbohydrate composition of horse beans (*Vicia faba*) of different origins. *Cereal Chem.* 52, 125–138.

Chen, L., Zheng, Y., Dong, Z., Meng, F., Sun, X., Fan, X., et al. (2018). Soybean (*Glycine max*) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. *Mol. Genet. Genomics* 293, 401–415. doi: 10.1007/s00438-017-1393-2

Dam, S., Laursen, B. S., Ørnfelt, J. H., Jochimsen, B., Stærfeldt, H. H., Friis, C., et al. (2009). The proteome of seed development in the model legume lotus japonicus. *Plant Physiol.* 149, 1325–1340. doi: 10.1104/pp.108.133405

Davidson, N. M., Hawkins, A. D. K., and Oshlack, A. (2017). SuperTranscripts: A data driven reference for analysis and visualisation of transcriptomes. *Genome Biol.* 18, 1–10. doi: 10.1186/S13059-017-1284-1/FIGURES/4

Demonsais, L., Utz-Pugin, A., Loubéry, S., and Lopez-Molina, L. (2020). Identification of tannic cell walls at the outer surface of the endosperm upon Arabidopsis seed coat rupture. *Plant J.* 104, 567–580. doi: 10.1111/tpj.14994

Ellwood, S. R., Phan, H. T. T., Jordan, M., Hane, J., Torres, A. M., Avila, C. M., et al. (2008). Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with Lens culinaris. *BMC Genomics* 9, 380. doi: 10.1186/1471-2164-9-380

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/BIOINFORMATICS/BTW354

Ezeagu, I. E., Petzke, J. K., Metges, C. C., Akinsoyinu, A. O., and Ologhobo, A. D. (2002). Seed protein contents and nitrogen-to-protein conversion factors for some uncultivated tropical plant seeds. *Food Chem.* 78, 105–109. doi: 10.1016/S0308-8146(02)00105-X

FAOSTAT (2021) *Licens. CC BY-NC-SA 3.0 IGO*. Available at: https://www.fao.org/faostat/en/#compare (Accessed April 26, 2023).

Focks, N., and Benning, C. (1998). wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.* 118, 91–101. doi: 10.1104/PP.118.1.91

Gallardo, K., Firnhaber, C., Zuber, H., Héricher, D., Belghazi, M., Henry, C., et al. (2007). A combined proteome and transcriptome analysis of developing medicago truncatula seeds: evidence for metabolic specialization of maternal and filial tissues. *Mol. Cell. Proteomics* 6, 2165–2179. doi: 10.1074/MCP.M700171-MCP200

Gao, M. J., Lydiate, D. J., Li, X., Lui, H., Gjetvaj, B., Hegedus, D. D., et al. (2009). Repression of seed maturation genes by a trihelix transcriptional repressor in arabidopsis seedlings. *Plant Cell* 21, 54. doi: 10.1105/TPC.108.061309

Grimberg, Å., Carlsson, A. S., Marttila, S., Bhalerao, R., and Hofvander, P. (2015). Transcriptional transitions in Nicotiana benthamiana leaves upon induction of oil synthesis by WRINKLED1 homologs from diverse species and tissues. *BMC Plant Biol.* 15, 1–17. doi: 10.1186/S12870-015-0579-1/FIGURES/9

Grimberg, Å., Wilkinson, M., Snell, P., De Vos, R. P., González-Thuillier, I., Tawfike, A., et al. (2020). Transitions in wheat endosperm metabolism upon transcriptional induction of oil accumulation by oat endosperm WRINKLED1. *BMC Plant Biol.* 20, 235. doi: 10.1186/s12870-020-02438-9

Guo, B., Sun, L., Jiang, S., Ren, H., Sun, R., Wei, Z., et al. (2022). Soybean genetic resources contributing to sustainable protein production. *Theor. Appl. Genet.* 135, 4095–4121. doi: 10.1007/S00122-022-04222-9

Gutierrez, N., Avila, C. M., and Torres, A. M. (2020). The bHLH transcription factor VfTT8 underlies zt2, the locus determining zero tannin content in faba bean (*Vicia faba* L.). *Sci. Rep.* 10, 1–10. doi: 10.1038/s41598-020-71070-2

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/NPROT.2013.084

Halling, J., and Hagman, M. (2020). "Sortval i ekologisk odling 2020," in *Sortförsök 2015–2019 i höstvete, höstråg, höstrågvete, vårvete, vårkorn, havre, åkerböna och potatis*. Uppsala: Department of crop production ecology, Swedish University of Agricultural Sciences, *Rapport No 29*.

Haughn, G., and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. *Trends Plant Sci.* 10, 472–477. doi: 10.1016/j.tplants.2005.08.005

Heuzé, V., Tran, G., Delagarde, R., Lessire, M., and Lebas, F. (2021)Feedipedia—Animal Feed Resources Information System, a programme by INRAE, CIRAD, AFZ and FAO. In: *Feed. a Program. by INRAE, CIRAD, AFZ FAO., Faba bean (Vicia faba)*. Available at: https://www.feedipedia.org/node/4926 (Accessed July 22, 2022).

Jayakodi, M., Golicz, A. A., Kreplak, J., Fechete, L. I., Angra, D., Bednář, P., et al. (2023). The giant diploid faba genome unlocks variation in a global protein crop. *Nature* 615, 652–659. doi: 10.1038/s41586-023-05791-5

Jia, H., Suzuki, M., and Mccarty, D. R. (2014). Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. *Wiley Interdiscip. Rev. Dev. Biol.* 3, 135–145. doi: 10.1002/wdev.126

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., and Hattori, T. (2005). LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46, 399–406. doi: 10.1093/PCP/PCI048

Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* 49, D192–D200. doi: 10.1093/NAR/GKAA1047

Kang, Y., Li, M., Sinharoy, S., and Verdier, J. (2016). A snapshot of functional genetic studies in Medicago truncatula. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01175

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi: 10.1093/bioinformatics/bts611

Krishnan, H. B., and Coe, E. H. (2001). Seed storage proteins. *Encycl. Genet.* 1782–1787. doi: 10.1006/rwgn.2001.1714

Krotz, L., and Galotta, W. (2020). *Elemental Analysis: Combustion (Dumas) method for Nitrogen/Protein determination of animal feed reference materials*. Application Note; 42497. Massachusetts: Thermo Fisher Scientific.

Lalanne, D., Malabarba, J., Ly Vu, J., Hundertmark, M., Delahaie, J., Leprince, O., et al. (2021). Medicago abi3 splicing isoforms regulate the expression of different gene clusters to orchestrate seed maturation. *Plants* 10 (8), 1710. doi: 10.3390/plants10081710

Le, D. T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K., et al. (2011). Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Res.* 18, 263–276. doi: 10.1093/DNARES/DSR015

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/S13059-014-0550-8/FIGURES/9

Lybæk, R., and Hauggaard-Nielsen, H. (2019). "The use of faba-bean cropping as a sustainable and energy saving technology - A new protein self-sufficiency opportunity for European agriculture?," in *IOP Conference Series: Earth and Environmental Science*. (Seoul: IOP Publishing Ltd), 291. doi: 10.1088/1755-1315/291/1/012049

Ma, W., Kong, Q., Arondel, V., Kilaru, A., and Bates, P. D. (2013). WRINKLED1, A ubiquitous regulator in oil accumulating tissues from *arabidopsis* embryos to oil palm mesocarp. *PloS One* 8, 68887. doi: 10.1371/journal.pone.0068887

MacNeill, G. J., Mehrpouyan, S., Minow, M. A. A., Patterson, J. A., Tetlow, I. J., and Emes, M. J. (2017). Starch as a source, starch as a sink: the bifunctional role of starch in carbon allocation. *J. Exp. Bot.* 68, 4433–4453. doi: 10.1093/JXB/ERX291

Malovichko, Y. V., Shtark, O. Y., Vasileva, E. N., Nizhnikov, A. A., and Antonets, K. S. (2020). Transcriptomic Insights into Mechanisms of Early Seed Maturation in the Garden Pea (*Pisum sativum* L.). *Cells* 9 (3), 779. doi: 10.3390/cells9030779

Miray, R., Kazaz, S., To, A., and Baud, S. (2021). Molecular control of oil metabolism in the endosperm of seeds. *Int. J. Mol. Sci.* 22, 1–23. doi: 10.3390/ijms22041621

Multari, S., Stewart, D., and Russell, W. R. (2015). Potential of fava bean as future protein supply to partially replace meat intake in the human diet. *Compr. Rev. Food Sci. Food Saf.* 14, 511–522. doi: 10.1111/1541-4337.12146

O'Sullivan, D. M., and Angra, D. (2016). Advances in faba bean genetics and genomics. *Front. Genet.* 7. doi: 10.3389/fgene.2016.00150

Oud, P. S., Henderik, J. B. J., Huysmans, A. C. L. M., Pahlplatz, M. M. M., Hermkens, H. G., Tas, J., et al. (1984). The use of light green and organge II as quantitative protein stains, and their combination with the feulgen method for the simultaneous determination of protein and DNA. *Histochemistry* 80, 49–57. doi: 10.1007/BF00492771

Patrick, J. W., and Stoddard, F. L. (2010). Physiology of flowering and grain filling in faba bean. *F. Crop Res.* 115, 234–242. doi: 10.1016/j.fcr.2009.06.005

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Punia, S., Dhull, S. B., Sandhu, K. S., and Kaur, M. (2019). Faba bean (*Vicia faba*) starch: Structure, properties, and *in vitro* digestibility—A review. *Legume Sci.* 1, e18. doi: 10.1002/LEG3.18

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/NAR/GKS1219

Röös, E., Carlsson, G., Ferawati, F., Hefni, M., Stephan, A., Tidåker, P., et al. (2020). Less meat, more legumes: Prospects and challenges in the transition toward sustainable diets in Sweden. *Renew. Agric. Food Syst.* 35, 192–205. doi: 10.1017/S1742170518000443

Ruuska, S. A., Girke, T., Benning, C., and Ohlrogge, J. B. (2002). Contrapuntal networks of gene expression during arabidopsis seed filling. *Plant Cell* 14, 1191–1206. doi: 10.1105/TPC.000877

Schmid, K. M. (2021). "Chapter 4 - Lipid metabolism in plants," in *Biochemistry of Lipids, Lipoproteins and Membranes (Seventh Edition)*. Eds. N. D. Ridgway and R. S. McLeod (Elsevier), 121–159. doi: 10.1016/B978-0-12-824048-9.00011-01

Schneider, A., Aghamirzaie, D., Elmarakeby, H., Poudel, A. N., Koo, A. J., Heath, L. S., et al. (2016). Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing *Arabidopsis thaliana* embryos. *Plant J.* 85, 305–319. doi: 10.1111/TPJ.13106

Semba, R. D., Ramsing, R., Rahman, N., Kraemer, K., and Bloem, M. W. (2021). Legumes as a sustainable source of protein in human diets. *Glob. Food Sec.* 28. doi: 10.1016/J.GFS.2021.100520

Sital, J. S., Malhotra, J. S., Sharma, S., and Singh, S. (2011). Comparative studies on biochemical components in mung bean [*Vigna radiata* (L.) Wilczek] varieties cultivated in summer and Kharif seasons. *Indian J. Agric. Biochem.* 24, 68–72.

Song, Y., Wang, X. D., and Rose, R. J. (2017). Oil body biogenesis and biotechnology in legume seeds. *Plant Cell Rep.* 36, 1519–1532. doi: 10.1007/S00299-017-2201-5

Stitt, M., and Zeeman, S. C. (2012). Starch turnover: pathways, regulation and role in growth. *Curr. Opin. Plant Biol.* 15, 282–292. doi: 10.1016/J.PBI.2012.03.016

Sun, S., Yi, C., Ma, J., Wang, S., Peirats-Llobet, M., Lewsey, M. G., et al. (2020). Analysis of Spatio-Temporal Transcriptome Profiles of Soybean (*Glycine max*) Tissues during Early Seed Development. *Int. J. Mol. Sci.* 21, 1–21. doi: 10.3390/IJMS21207603

Troncoso-Ponce, M. A., Kilaru, A., Cao, X., Durrett, T. P., Fan, J., Jensen, J. K., et al. (2011). Comparative deep transcriptional profiling of four developing oilseeds. *Plant J.* 68, 1014–1027. doi: 10.1111/j.1365-313X.2011.04751.x

Verdier, J., Kakar, K., Gallardo, K., Le Signor, C., Aubert, G., Schlereth, A., et al. (2008). Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol. Biol.* 67, 567–580. doi: 10.1007/s11103-008-9320-x

Warsame, A. O., Michael, N., O'Sullivan, D. M., and Tosi, P. (2020). Identification and quantification of major faba bean seed proteins. *J. Agric. Food Chem.* 68, 8535–8544. doi: 10.1021/acs.jafc.0c02927

Warsame, A. O., Michael, N., O'Sullivan, D. M., and Tosi, P. (2022). Seed development and protein accumulation patterns in faba bean (*Vicia faba*, L.). *J. Agric. Food Chem.* 70, 9295–9304. doi: 10.1021/acs.jafc.2c02061

Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi: 10.1146/annurev.arplant.56.032604.144201

Wu, T., Hu, E., Xu, S., Chen, M., Gou, P., and Dai, Z. (2021). clusterProfiler 4.0:Ú universal enrichment tool for interpreting omics data. *Innovation* 2, 100141. doi: 10.1016/j.xinn.2021.100141

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R packageßor comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, H., Hu, Z., Yang, Y., Liu, X., Lv, H., Song, B. H., et al. (2021). Transcriptome profiling reveals the spatial-temporal dynamics of gene expression essential for soybean seed development. *BMC Genomics* 22, 1–13. doi: 10.1186/s12864-021-07783-z

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014

II

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 March 2024
DOI 10.3389/fpls.2024.1348014

OPEN ACCESS

EDITED BY
Yi-Hong Wang,
University of Louisiana at Lafayette,
United States

REVIEWED BY
Cathrine Kiel Skovbjerg,
Nordic Seed A/S, Denmark
Ahmed Sallam,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
Fouad Maalouf,
International Center for Agricultural Research in the Dry Areas, Lebanon
Kenani Chiwina,
University of Arkansas, United States

*CORRESPONDENCE
Hannah Ohm
hannah.ohm@slu.se
Åsa Grimberg
asa.grimberg@slu.se

RECEIVED 01 December 2023
ACCEPTED 21 February 2024
PUBLISHED 06 March 2024

CITATION
Ohm H, Åstrand J, Ceplitis A, Bengtsson D, Hammenhag C, Chawade A and Grimberg Å (2024) Novel SNP markers for flowering and seed quality traits in faba bean (*Vicia faba* L.): characterization and GWAS of a diversity panel.
Front. Plant Sci. 15:1348014.
doi: 10.3389/fpls.2024.1348014

Novel SNP markers for flowering and seed quality traits in faba bean (*Vicia faba* L.): characterization and GWAS of a diversity panel

Hannah Ohm[1*], Johanna Åstrand[1,2], Alf Ceplitis[2], Diana Bengtsson[2], Cecilia Hammenhag[1], Aakash Chawade[1] and Åsa Grimberg[1*]

[1]Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Lomma, Sweden, [2]Lantmännen Agriculture, Plant Breeding, Svalöv, Sweden

Faba bean (*Vicia faba* L.) is a legume crop grown in diverse climates worldwide. It has a high potential for increased cultivation to meet the need for more plant-based proteins in human diets, a prerequisite for a more sustainable food production system. Characterization of diversity panels of crops can identify variation in and genetic markers for target traits of interest for plant breeding. In this work, we collected a diversity panel of 220 accessions of faba bean from around the world consisting of gene bank material and commercially available cultivars. The aims of this study were to quantify the phenotypic diversity in target traits to analyze the impact of breeding on these traits, and to identify genetic markers associated with traits through a genome-wide association study (GWAS). Characterization under field conditions at Nordic latitude across two years revealed a large genotypic variation and high broad-sense heritability for eleven agronomic and seed quality traits. Pairwise correlations showed that seed yield was positively correlated to plant height, number of seeds per plant, and days to maturity. Further, susceptibility to bean weevil damage was significantly higher for early flowering accessions and accessions with larger seeds. In this study, no yield penalty was found for higher seed protein content, but protein content was negatively correlated to starch content. Our results showed that while breeding advances in faba bean germplasm have resulted in increased yields and number of seeds per plant, they have also led to a selection pressure towards delayed onset of flowering and maturity. DArTseq genotyping identified 6,606 single nucleotide polymorphisms (SNPs) by alignment to the faba bean reference genome. These SNPs were used in a GWAS, revealing 51 novel SNP markers significantly associated with ten of the assessed traits. Three

markers for days to flowering were found in predicted genes encoding proteins for which homologs in other plant species regulate flowering. Altogether, this work enriches the growing pool of phenotypic and genotypic data on faba bean as a valuable resource for developing efficient breeding strategies to expand crop cultivation.

KEYWORDS

Vicia faba (faba bean), diversity panel, GWAS (genome wide association study), DArT-seq, field trial, SNPs (single nucleotide polymorphism), flowering

1 Introduction

Given the mounting environmental concerns regarding climate change and planetary boundaries, developing a more sustainable agriculture and food production system is an urgent global goal (Steffen et al., 2015). Legumes play an important role in promoting sustainable agriculture due to their agronomic, environmental, and nutritional benefits (Voisin et al., 2014; Rubiales et al., 2021). Cultivation of legumes is beneficial in crop rotations due to their capacity to fix atmospheric nitrogen through symbiosis with soil bacteria, replacing the need for synthetic nitrogen fertilizer. In addition, legumes break the disease cycles in cereal-dense agricultures (Rubiales et al., 2021), promoting reduced pesticide use. Due to their nutritional benefits and high protein content, legumes could replace animal-based protein in human diets and are therefore considered a prerequisite for transitioning to a more sustainable food system (Willett et al., 2019). However, only a small number of pulse crops are currently contributing to the world's food production among which soybean (*Glycine max*) accounts for 80% of the total legume production (*FAOSTAT, 2021*). To promote the cultivation and consumption of a diverse array of legumes, it is essential to provide growers with cultivars that exhibit high and stable yields and seed qualities valued by food producers. Therefore, there is an urgent demand for increased plant breeding efforts on pulse crops.

Faba bean [*Vicia faba* L. (Fabaceae)] is a leguminous crop characterized by seeds with a relatively high protein content (30%), and holds a high potential for increased cultivation and use in food. However, current challenges to its cultivation and use are unstable yields, disease susceptibility, sensitivity to drought, antinutritional compounds such as tannins, as well as convicin and vicin that cause favism in individuals with a specific hereditary disease (Maalouf et al., 2018; Björnsdotter et al., 2021). Further, it is a partially allogamous species with outcrossing rates between 4% and 84% (Ellwood et al., 2008). While no wild ancestors have yet been found, over 38,000 accessions of *V. faba* are available in genebanks worldwide, represented in collections by cultivated forms only (Duc et al., 2010). Even though a wide diversity of plant and seed phenotypes has been reported (Maalouf et al., 2018), the scarce phenotypic and genotypic characterization of germplasm collections limits the exploitation of available diversity as a resource for breeding.

Faba bean has one of the largest genomes of the diploid crops, with approximately 13 Gb distributed across six pairs of chromosomes, and possibly around 85% composed of repetitive DNA (Khazaei et al., 2021). Detailed genetic and physical maps of markers and quantitative trait loci (QTL) have previously not been available for faba bean. However, several genetic consensus maps based on biparental populations and single nucleotide polymorphism (SNP) markers identified from transcriptome data, together with genotyping-by-sequencing (GBS) markers mapped to publicly available faba bean genomic and transcriptomic sequences, have been developed lately (Webb et al., 2016; Carrillo-Perdomo et al., 2020; Wang et al., 2021; Abou-Khater et al., 2022; Maalouf et al., 2022; Li et al., 2023; Zhao et al., 2023). Recently, a significant milestone was achieved with the publication of the first reference genome of faba bean, providing a valuable resource for further genetic analysis (Jayakodi et al., 2023).

Genome-wide association study (GWAS) has become an important tool for geneticists in identifying genomic loci governing target traits in plants and is based on the association of genotypic and phenotypic data (Torkamaneh and Belzile, 2022). Through GWAS on faba bean, genetic markers have been associated with agronomic traits such as heat stress and herbicide tolerance, plant architecture (height, branching, and flower/pod placement), time to flowering, and seed yield (Abou-Khater et al., 2022; Maalouf et al., 2022; Karaköy et al., 2023; Li et al., 2023). Several GWAS on faba bean with a focus on seed quality aspects were also recently published, making use of the available reference genome to localize markers. Jayakodi et al. (2023) identified genetic markers for seed size based on a 90K SPET genotyping assay on a diversity collection of 197 accessions, and Skovbjerg et al. (2023) genotyped a 7-parent MAGIC population with a 21,345 SNP array which allowed for the identification of 238 markers for agronomic and seed quality traits. Further, Zhao et al. (2023) genotyped 121 individuals from an F2 population using a 130K SNP chip which was developed from RNA-seq data on flowering and leaf tissues (Wang et al., 2021), which allowed for the identification of markers associated with 65 seed traits, including seed quality.

To date, only a limited number of GWAS on faba bean have utilized diversity panels, and none of them have considered both

agronomic and seed quality traits. In this study, we assembled a diversity panel of 220 accessions of faba bean originating from diverse geographical regions, sourced from gene bank material and commercially available varieties. The aims of this study were to i) characterize this diversity panel for agronomic- and seed quality traits in field trials at Northern latitudes, ii) analyze the effects of selection pressure through breeding on faba bean to date, and iii) to identify genetic markers (SNPs) associated with target traits through GWAS.

2 Materials and methods

2.1 Plant material

The diversity panel in this study comprised 220 faba bean accessions selected to represent a wide variation of geographical origin, flower and seed color, seed size, plant height, tannin and convicine/vicine content, and breeding advancement status. The sources of seeds were genebanks (Nordic Genetic Resource Centre (SWE054, https://www.nordgen.org/en/), Genebank of Leibniz Institute of Plant Genetics and Crop Plant Research (DEU146, https://www.ipk-gatersleben.de/), and National Plant Germplasm System of the United States Department of Agriculture (USA022, https://npgsweb.ars-grin.gov/gringlobal/search) and commercially available varieties from breeders in Europe. The country of origin for each accession and link to the source of seeds are found in Supplementary Datasheet 1. A heat map of the distribution of the accessions based on origin was generated using the ggplot2 package in R (Wickham, 2016).

Based on passport data, the 220 accessions from the diversity panel were classified into four categories with increasing level of advancement. The categories were *cultivated* (plant material with scarce information annotated by gene banks as cultivated, unknown, or 'wild'), *landraces* (plant material annotated either as landrace by gene banks or known heirloom cultivars, including accessions from a seed preservation and collection program in Sweden), *advanced* (plant material that has undergone some level of advancement through either research or breeding but not registered as a variety in the European Community Plant Variety Office database (CPVO) as per 2022-10-14) and *varieties* (varieties registered in the CPVO database as per 2022-10-14, including both agricultural and vegetable varieties). The term *cultivated* is considered as the least developed plant material.

Seeds from the 220 accessions were amplified in the field or greenhouse during 2020. The field site for seed amplification was situated at 55.90_N, 13.09_E, where the accessions with small- to medium-sized seeds were sown by machine, and accessions with larger seeds were sown manually. Sowing was conducted in mid-April and manual harvesting occurred at maturity from late July to August. Seed amplification in the greenhouse was conducted between November 2020 to February 2021 by growing single plants in 7.5 L pots in soil fertilized with 3M Plus Basacote (Compo Expert, Muenster, Germany). The temperatures were 21° C (day) and 18°C (night), at a relative humidity of 70%, with a 16 h photoperiod, supplemented with Son-T PIA 400 W sodium lamps (Philips, Amsterdam, Netherlands) when the natural light fell below 200 $\mu mol \cdot m^{-2} s^{-1}$ photosynthetically active radiation. Plants were isolated during flowering either by spatial separation (in the greenhouse) or with perforated plastic isolation bags (in the field) to prevent cross-pollination between accessions. Pods were manually harvested, threshed and stored at 4°C.

2.2 Field trial design

The field experiments were carried out at the SITES Research Station Lönnstorp, SLU, in Alnarp (55.65_ N, 13.06_ E) spanning two consecutive growing seasons in 2021 and 2022, in an alpha lattice design with 11 blocks and two replicates. The field was fertilized prior to sowing (300 kg/ha Pk11-21 in 2021, 350 kg/ha Pk11-21 in 2022). Fifty seeds were hand-sown in each plot à 1 m x 0.75 m, resulting in a density of 66 seeds/m^2 and an average seed distance of approximately 15 cm. Seeds used in the 2021 field trial derived from seed amplification conducted during 2020, whereas seeds for the 2022 field trial derived from the harvest of the 2021 field trial. During both trials, the field was covered with a fiber cloth until plant establishment, to protect the germinated seedlings from birds. The field was watered as needed (four times early in the growth season of 2021 to secure seedling establishment, none in 2022), manually weeded between plots, and sprayed against weeds (Corum® once in 2021, Corum® and Dash® once in 2022) and aphids (MAVRIK® AQUAFLO once in 2021, Teppeki® once in 2022). For detailed weather data, soil conditions, and further information on field management, see Supplementary Datasheet 1.

2.3 Phenotyping of agronomic and seed quality traits

Field phenotypic data were collected for the following parameters: establishment (number of established plants per plot ~35 days after sowing), plant height (average height of five randomly selected plants/plot measured at ~75 days after sowing), days from sowing to flowering (when 50% or more of the established plants in a plot had at least one open flower), days from sowing to maturity (when 50% or more of the established plants in a plot had filled, dry and brown/black pods). The damage caused by the broad bean weevil (*Bruchus rufimanus* Boh.) and its parasitoids was assessed after harvest by classifying 100 randomly selected seeds as healthy or infested seeds as described by Carrillo-Perdomo et al. (2019). The testa color of the seeds was determined by classifying them into four color categories (dark, red, green, and light). The gradients of reference colors for each category were defined by digitally sampling photos of seeds with a color picker tool using Affinity Photo software (Supplementary Figure 1). Seed dimensions (length, width, and area) and thousand-grain weight (TGW) were determined using 100-200 seeds with the seed analyzer MARViN ProLine I (Marvitech, Germany). The yield parameter expressed as gram seeds per plant (YIELD) was determined by dividing the total weight of harvested seeds from a plot (g) by the number of established plants in that plot. The total number of seeds

per plant (SEEDS) was determined by dividing the total weight of harvested seeds from a plot by its TGW, then multiplying by 1000, and finally dividing by the number of established plants in that plot.

2.4 Seed protein and starch analysis

For protein and starch analysis, approximately 5 g seeds (at least 10 seeds) from each of the 220 faba bean accessions from the 2021 field trial were ground into flour using a centrifugal mill at 10,000 rpm and passed through a 0.5 mm sieve (ZM 200, Retsch GmbH Haan, Germany) and subsequently freeze-dried for 48 h at -60°C. Raw protein content was analyzed using the Dumas method (N×6.25) on 0.5 g flour (Eurofins Food & Feed Testing Sweden AB, Lidköping, Sweden).

Total starch content was determined using the Megazyme Total Starch Assay Kit K-TSTA (Megazyme, Bray, Ireland) on 80 mg flour. The assay is based on starch degradation with α-amylase and amyloglucosidase, followed by colorimetric determination of the released glucose. To ensure accurate starch determination and minimize interference, any potential presence of free sugars in the flour was extracted with ethanol washes prior to the starch analysis. Thereafter, the recommended procedure was followed, with the exception of a downscaled enzymatic colorimetric conversion that facilitated more convenient handling of the samples. The absorbance was measured at 510 nm on a microplate spectrophotometer (Multiskan GO, ThermoFisher Scientific, MA, USA).

2.5 Statistical analysis of the phenotypic data

The best linear unbiased estimates (BLUE) values, broad-sense heritability, and variance components of each trait were obtained using the META-R software with adjustments following the alpha lattice design (Alvarado et al., 2020). The model used for the calculation of the BLUE values was:

$$Y_{ijkl} = \mu + Loc_i + Rep_j(Loc_i) + Block_k(Loc_i Rep_j) + Gen_l + Loc_i \times Gen_l + \varepsilon_{ijkl}$$

where Y_{ijkl} is the trait of interest, μ is the overall mean effect, Loc_i is the effect of the ith location, $Rep_j(Loc_i)$ is the effect of the jth replicate in the ith location, $Block_k(Loc_i Rep_j)$ is the effect of the kth block within the ith location and the jth replicate, Gen_l is the effect of the lth genotype, the $Loc_i \times Gen_l$ is the effect of the environment × genotype interaction, and ε_{ijkl} is the effect of the error associated with the ith location, the jth replication, the kth block, and the lth genotype, as specified in Alvarado et al. (2020). The BLUE values for each trait of all accessions were obtained from two replicates from each year (with the exception of protein content, starch content, and bean weevil damage, for which only one year of data was available) by assuming fixed genotypic and random environmental effects. Broad-sense heritability was calculated using:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{nLoc} + \frac{\sigma_\varepsilon^2}{nLoc \times nRep}}$$

where σ_g^2, σ_ε^2, and σ_{ge}^2 are the genotype, error, and genotype by environment interaction variance components, respectively, nLoc is the number of environments, and nRep is the number of replicates, as in Alvarado et al. (2020). The BLUE value of each accession was used for linear pairwise correlations for each trait and were assessed using Pearson's correlation coefficient, with an ANOVA t-test where p< 0.001: ***, p< 0.01: **, p< 0.05: *. Differences between the groups, defined by the different levels of advancement, were estimated using a pairwise t-test with Bonferroni adjustment.

2.6 DNA extraction and DArTseq genotyping

Leaf tissue for DNA extraction was sampled from greenhouse-grown plants in March 2022. Seeds from each accession in the diversity panel (from isolated plants 2020, see above) were sown in 2 L pots filled with soil (50% peat, pH 5.5-6.5, added per m^3 soil: 5.5 kg lime, NPK 11-5-18 kg, 200 g micronutrients and 100 g iron). Greenhouse parameters were as described above. Two weeks after germination, a single true leaf from five plants per accession was sampled and put in a petri dish on ice. Two 3 mm diameter discs were punched from each of the five leaves, resulting in a total of ten discs per accession. The discs were pooled and placed in a single well of a 96-well plate on ice. For eight of the accessions less than five plants were available, and the sampling was therefore evenly distributed among the available plants to obtain ten discs. The puncher and punching mat were carefully cleaned between sampling of each accession, to prevent cross-contamination. The leaf tissues were freeze-dried for 40 h at -60°C and then kept at room temperature until DNA extraction. In total, 187 of the 220 accessions from the diversity panel were sampled. DNA extraction was performed by Intertek ScanBi Diagnostics (Alnarp, Sweden) using sbeadex™ plant DNA extraction kit (Biosearch Technologies, Hoddesdon, United Kingdom). A high DNA quality was confirmed by checking DNA integrity using agarose (1%) gel electrophoresis. The samples were genotyped by DArTseq (Diversity Array Technologies, Canberra, Australia), which is a genome complexity reduction-based sequencing using restriction enzymes. The sequencing depth used was 800 000 counts/sample.

2.7 Marker filtering and sequence mapping to the reference genome

Through DArTseq 19,770 SNP markers were identified, and their flanking sequences were aligned using BLAST+ to the reference genome of *Vicia faba* 'Hedin/2' (Jayakodi et al., 2023) to determine their respective positions within the genome. The aligned markers were filtered for >95% replication average and a call rate >50%. Imputation of the missing data in the remaining markers was conducted in TASSEL 5.0 (Bradbury et al., 2007) using the five closest neighboring accessions and Eucledian distance. A subsequent filtering step was applied to select markers with only

one aligned position in the *V. faba* ('Hedin/2') reference genome and with an aligned sequence having an E-value $>10^{-6}$, resulting in the identification of 6,606 markers in the *V. faba* genome. Phylogenetic distance between the accessions was estimated using neighboring clustering in TASSEL 5.0 (Bradbury et al., 2007) and visualized with Interactive tree of life (iTOL) (Letunic and Bork, 2021). SNP density was visualized using the rMVP R package (Yin et al., 2021). Linkage disequilibrium was calculated with the imputed genotype data in TASSEL 5.0 (Bradbury et al., 2007) and visualized in R 4.2.2 using script written by Mohsin (2021) based on Remington et al. (2001).

2.8 Genome-wide association study

A GWAS was conducted for each trait using the Genome Association and Prediction Integrated Tool (GAPIT) 3.0 function (Wang and Zhang, 2021) implemented in R 4.2. Principal component analysis (PCA) and the VanRaden kinship matrix were conducted using the GAPIT 3.0 function to determine any underlying population structure. The GWAS was conducted for each trait separately using the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) model implemented in the GAPIT 3.0 function, with zero principal components and a minor allele frequency (MAF) threshold set to 0.01. Predicted genes in which the associated markers were localized were annotated based on the faba bean reference genome, as well as on the highest sequence similarity of legume species using blastX at NCBI using default settings (BLAST).

3 Results

3.1 The diversity panel

The diversity panel had a broad geographical distribution with representatives from all continents, with a majority of the accessions originating from northern and central Europe (34% and 23%, respectively) and North Asia (14%) (Figure 1A). Based on their origin and passport data (Supplementary Datasheet 1), the 220 accessions from the diversity panel were classified into four categories with increasing degree of advancement (Figure 1B). All four categories were represented by approximately 60 accessions, except for the least advanced category, which comprised only 30 accessions. The two most advanced categories were primarily composed of accessions of European origin, while the category "landrace" included accessions from all geographical regions. The category "cultivated" on the other hand, were predominantly composed of non-European material.

3.2 Phenotypic diversity and correlations

The phenotypic raw data of the 220 accessions in the diversity panel displayed a broad range of values for several key agronomic traits (Supplementary Datasheet 1). Flower colors of brown, white, purple, red, and variegated (i.e., white petals with a black wing melanin spot), as well as various seed shapes and testa colors, were represented (Figures 2A–C). In the field trials during 2021 and 2022, the average establishment of the 50 sown seeds was 82-91% approximately 30 days after sowing. The phenotypic raw data was used to estimate the BLUE values for each year (Supplementary Figure 2) and for both years (Figure 3). Phenotype metrics of BLUE values based on both years' phenotype data are shown in Table 1. All measured traits showed broad-sense heritability values above 0.6 and with the majority substantially exceeding 0.8 (Table 1). Time from sowing to flowering in the diversity panel varied between 52-70 days (Figure 3A). While most of the accessions flowered between 55-62 days after sowing, six accessions consistently exhibited very early flowering (<55 days) and 17 accessions consistently displayed very late flowering (>66 days) across both years. Time from sowing to maturity varied between 98-114 days, with most accessions maturing at day 108 (Figure 3B). At the time point of growth stabilization (~77 days after sowing), the plant height showed a

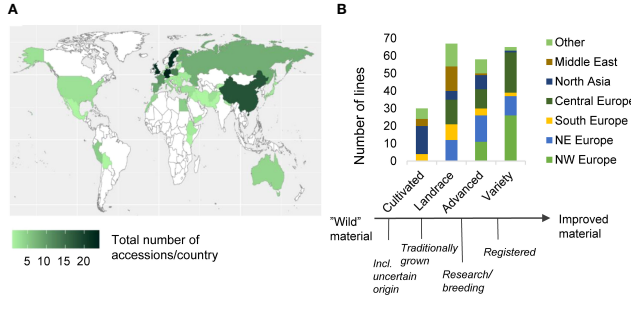


FIGURE 1
Overview of the diversity panels' geographical origin and breeding advancement status, and genetic relatedness. Distribution of the accessions' geographical origin per country **(A)**. Geographical distribution of the accessions per country, sorted on their breeding advancement **(B)**. For the definitions of the categories, see Materials & Methods.

distribution range between 60-106 cm, but one dwarf accession only attained 33 cm (Figure 3C). Seed yield (g per plant) showed a substantial variation between accessions (4-17 g/plant), with normal distribution (Figure 3D). Twenty-one of the highest-yielding accessions (>12 g/plant) performed consistently across both years. Thousand-grain weight (TGW) varied between small, pea-sized accessions with values below 200 g up to large-seeded accessions at 1200 g (Figures 2B and 3E), which was also reflected in the variation in seed size (i.e., seed area, Figure 3G). The number of seeds per plant varied from 7 to 37, with a few outliers reaching as high as 56 (Figure 3F). Seed protein and starch content in the diversity panel exhibited substantial variation, with values ranging from approximately 25 to 37% for both traits (Figures 3H, I). The level of bean weevil damage was generally high and showed a variation between 45 to 100% infested seeds in the accessions (Figure 3J).

3.3 Pairwise correlation coefficient analysis of agronomic traits

To examine phenotypic pairwise correlations among the different traits characterized in the diversity panel, a Pearson's correlation coefficient matrix was computed [Figure 4 and Supplementary Table 1 for correlation coefficient values (r)]. Interestingly, several traits showed strong correlations, with all of the mentioned correlations being highly significant at $p < 0.01$ or lower. The positive correlation found between TGW and seed size confirms the intuitive relationship of larger seeds having a higher weight (r = 0.99). An inverse correlation was observed for seed size and number of seeds per plant (r = -0.77), indicating a common biological trade-off where larger seeds are associated with fewer seeds per plant. Yield was positively correlated with days to maturity (r = 0.53), plant height (r = 0.49), number of seeds per plant (r = 0.40), and moderately with TGW (r = 0.23).

Days to flowering was positively correlated with days to maturity (r = 0.63), indicating a clear relationship where plants that flowered later also matured later. Later flowering plant material was associated with a higher number of seeds per plant (r = 0.50) and taller plants (r = 0.37). Interestingly, the data indicated that yield was independent of days to flowering but positively associated with days to maturation. This indicates that a later maturation is beneficial for a higher yield, whereas later time to flowering is not. Furthermore, days to flowering was negatively correlated with TGW (r = -0.50) and seed size (r = -0.53), indicating that accessions with later flowering tend to have smaller seeds.

Regarding nutritional seed quality traits, a negative correlation was observed between protein content and starch content (r = -0.38). Furthermore, both starch content and protein content were negatively correlated with TGW (r = -0.24 and r = -0.27, respectively) and seed size (r = -0.30 and r = -0.27, respectively). These observations imply that smaller seeds tend to have a higher content of both protein and starch, as compared to larger seeds. Consequently, it can be inferred that the content of other seed constituents is higher in larger seeds. Interestingly, protein content did not show any correlations with yield but a slight positive correlation with number of seeds (r = 0.27). For starch content, there was only a weak positive correlation indicated with yield (r = 0.17, at $p < 0.05$) but a clear correlation with plant height (r = 0.48).

Bean weevil damage of seeds exhibited a strong negative correlation with days to flowering (r = -0.68) and maturity (r = -0.60), indicating that plants that flowered and matured earlier were more prone to have a higher degree of infested seeds. Correlations were observed for weevil damage with both seed size (r = 0.61) and TGW (r = 0.55), which implied that accessions with larger seeds were more infested. Correlations also showed that

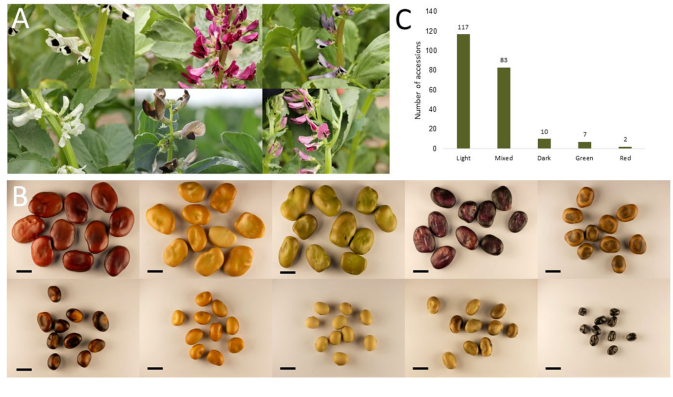


FIGURE 2
Examples of variation in flower color **(A)** and seed color, shape, and size **(B)** present in the diversity panel used in this study. Graph **(C)** shows the grouping of the whole diversity panel into distinct seed color categories shown in Supplementary Figure 1, with the panels below showing the color scales included in every category. Scale bar in B is 1 cm.

10.3389/fpls.2024.1348014

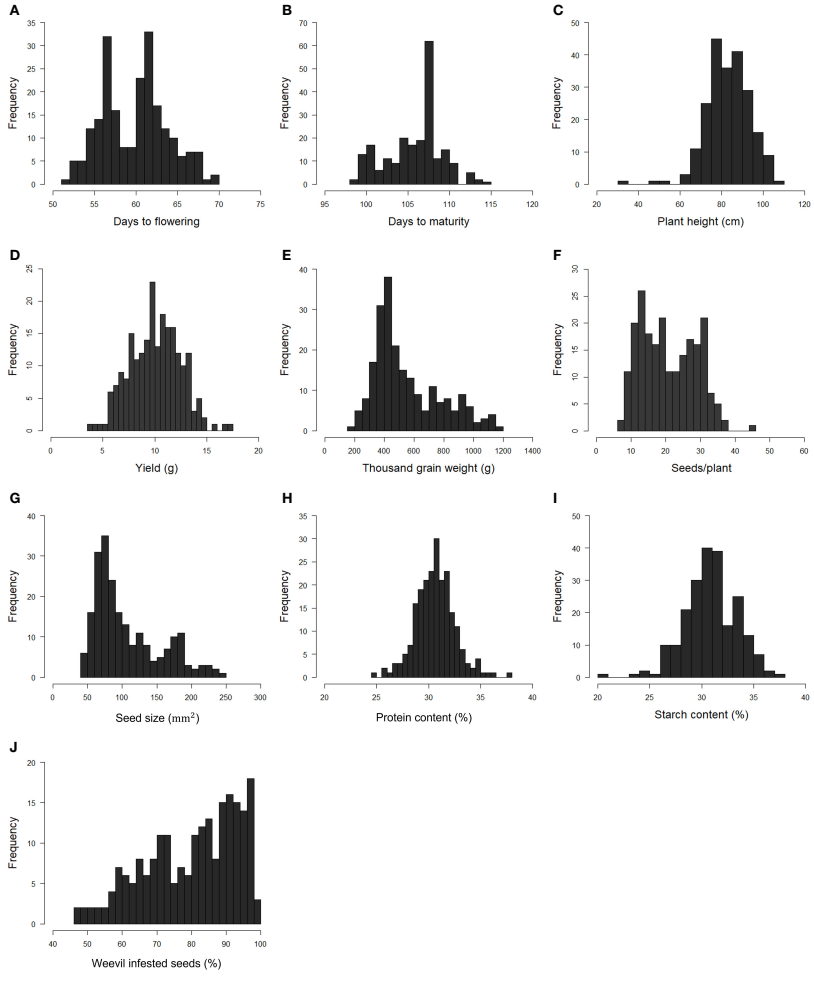


FIGURE 3
Histograms showing the frequency distribution of Best Linear Unbiased Estimates (BLUE) values of each trait evaluated in the faba bean diversity panel in field trials across two years 2021 and 2022*. Days to flowering **(A)**. Days to maturity **(B)**. Plant height **(C)**. Seed yield per plant **(D)**. Thousand-grain weight **(E)**. Number of seeds per plant **(F)**. Seed size **(G)**. Seed protein content by dry weight **(H)**. Seed starch content by dry weight **(I)**. Proportion of seeds infested by bean weevil **(J)**. *Data for bean weevil, protein content, and starch content are only for the 2021-year field trial.

TABLE 1 Phenotype metrics, broad-sense heritability (H^2), and coefficient of variation (CV) for each trait characterized in the diversity panel consisting of 220 accessions.

	Min	Max	Mean	H^2	CV	Accession with min value	Accession with max value
HEIGHT (cm)	33.60	106.00	83.16	0.89	6.56	Dwarf Ö53	Ashleigh
DTF (days)	52.0	69.5	59.8	0.94	2.28	Felix	CHA CHA
DTM (days)	98.25	114.00	105.85	0.78	2.00	ATC 63752	Banquise

(Continued)

10.3389/fpls.2024.1348014

TABLE 1 Continued

	Min	Max	Mean	H^2	CV	Accession with min value	Accession with max value
TGW (g)	155.20	1161.96	549.87	0.97	10.10	Mikko	FAB 590
YIELD (g/plant)	3.91	17.08	10.05	0.65	22.38	Mikko	Karmazyn
SIZE (mm^2)	41.60	248.60	107.10	0.98	7.59	Mikko	Super Aquadulce
SEED WIDTH (mm)	6.43	15.50	10.03	0.98	3.60	Mikko	Super Aquadulce
SEED LENGTH (mm)	8.35	20.93	13.13	0.98	3.62	Mikko	Grebo
WEEVIL (% infested seeds)	47.4	99.3	80.06	0.90	29.54	COLUMBA	Syria local small
SEEDS (no/plant)	6.69	44.13	20.72	0.88	23.59	Super Aquadulce	Nanaux
PROTEIN (%)	25.05	38.05	30.64	0.71	4.73	ATC 63759	Dwarf Ö53
STARCH (%)	20.93	37.84	30.86	0.66	6.63	Dwarf Ö53	Ticol HÖG

Min, max, and mean for each trait are given as the Best Linear Unbiased Estimates (BLUE) values across years. The accessions with the min and max values of each trait are given to the right, for phenotypic data of each accession in the panel see Supplementary Datasheet 1. HEIGHT, plant height (cm); DTF, days to flowering; DTM, days to maturity; TGW, thousand grain weight (g); YIELD, seed weight per plant (g); SIZE, seed size as area (mm2); WEEV, proportion of seeds infested by bean weevil (%); SEEDS, number of seeds per plant; PROTEIN, seed protein content by dry weight (%); STARCH, seed starch content by dry weight (%). STARCH, seed starch content by dry weight (%). Data for WEEVIL, PROTEIN, and STARCH is from 2021 only.

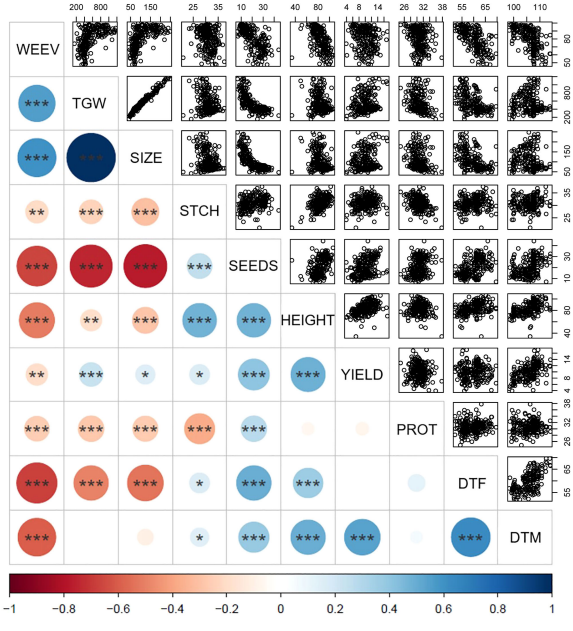


FIGURE 4
Pairwise correlations between traits characterized in the diversity panel based on phenotypic values. The color intensity and circle size represent the strength of Pearson's correlation coefficient with larger circle sizes indicating stronger correlation and the color scale indicating negative correlations in red and positive correlations in blue according to the scale at the bottom. Number of stars indicates the significance level based on ANOVA t-test with $p < 0.001$: ***, $p < 0.01$: **, $p < 0.05$: *, and $p > 0.05$: no star, and correlation values for all traits are found in Supplementary Table 1. The scatter plots in the upper panel show all correlated data points. WEEV, proportion of seeds infested by bean weevil (%); TGW, thousand grain weight (g); SIZE, seed size as area (mm^2); DTF, days to flowering; DTM, days to maturity; HEIGHT, plant height (cm); YIELD, seed weight per plant (g); SEEDS, number of seeds per plant, PROT, seed protein content by dry weight (%), STCH, seed starch content by dry weight (%).

10.3389/fpls.2024.1348014

increased bean weevil damage was significantly associated with fewer seeds per plant (r = -0.66) and shorter plant height (r = -0.52). Interestingly, the extent of bean weevil damage showed only a slight correlation with lower yields (r = -0.19). A slight negative correlation was found between infested seeds and starch (r = -0.21) and protein levels (r = -0.26).

3.4 Distribution of traits depending on advancement status

The distribution of the agronomic traits, grouped based on their breeding status, is visualized in the violin plots of Figure 5. Significant differences were observed between more improved and less improved material, with the former showing an increase in days to flowering and days to maturity. In addition, compared to its less improved counterpart, the more improved material exhibited higher yield, more seeds per plant, and taller plants. Although there was a trend towards decreasing TGW and seed size for more advanced material, these differences did not reach statistical significance. Notably, the level of bean weevil damage was lower on the improved plant material compared to the less improved counterpart. Higher bean weevil damage on the less improved material could potentially be linked to the earlier flowering observed in this category, which is negatively correlated to bean weevil damage (Figure 4). However, even though the level of damage on the improved material was lower, more variation was observed within this group as compared to the least improved material. No significant patterns of change for seed protein and starch content through breeding could be observed.

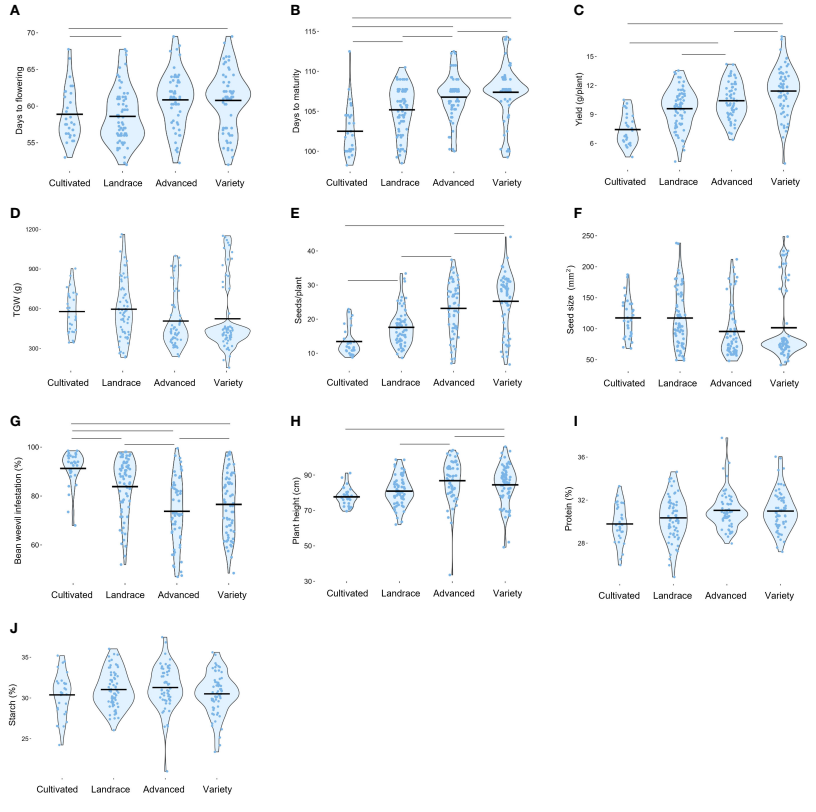


FIGURE 5
Trait distributions are based on the advancement status of the accessions. Data shown are the Best linear unbiased estimates (BLUE) values of each trait evaluated in the faba bean diversity panel in field trials across two years 2021 and 2022*. Each dot represents one accession. The bold lines in each violin plot are the mean values. Days to flowering **(A)**. Days to maturity **(B)**. Seed yield per plant **(C)**. Thousand-grain weight **(D)**. Number of seeds per plant **(E)**. Seed size as seed area **(F)**. Proportion of seeds infested by bean weevil **(G)**. Plant height **(H)**. Seed protein content by dry weight **(I)**. Seed starch content by dry weight **(J)**. *Data for bean weevil, protein content, and starch content is only for 2021 years' field trial. Lines between groups indicate a significance level of at least p > 0.05. Note that sample sizes differ between groups and that non-significant pairs may be due to differences in statistical power.

3.5 Marker filtration and mapping to reference genome

To identify markers in the faba bean genome for genetic analysis, 187 of the accessions in the diversity panel were genotyped using the genotyping by sequencing method DArTSeq, resulting in the identification of 19,770 single nucleotide polymorphism (SNPs) with a replication average of 97.4%. However, the average call rate of 56.6% indicated a large proportion (ca. 40%) of missing data. After filtering (for a replication average of >95% and a call rate of >50%) 8,478 SNPs remained. These markers had an average of 29.6% missing data which was estimated through imputation using the five nearest neighbors. To infer a position of each SNP in the *V. faba* genome, the marker sequences were mapped to the 'Hedin/2' reference genome (Jayakodi et al., 2023), resulting in 6,606 markers with a single mapped position. These remaining markers were evenly spread throughout the genome and over the chromosomes (Figure 6A) with an average SNP density of 0.5 SNP/Mbp. See Supplementary Datasheet 2 for a full table of SNPs and sequencing data and Supplementary Figure 3 for a linkage disequilibrium (LD) decay plot. The decay plot showed a low LD between markers, probably due to the large size of the faba bean genome and the low marker density.

V. faba is a partially outcrossing species, and the diversity panel was composed of accessions with varying levels of inbreeding. To obtain the best approximation of the genotype for each accession, five plants per accession were pooled before sequencing. This resulted in an average marker heterozygosity of 13.4%. In addition, there was a high frequency of rare alleles, with an average MAF of 8.5% throughout the population (Figure 6B). The 6,606 markers were used to estimate kinship in the population using the VanRaden method which gives a matrix where the value 0 means no genetic relatedness and 2 means complete relatedness i.e. genetically identical (Figure 6C). Due to the broad diversity of the panel, the relatedness amongst accessions was found to be very low. A principal component analysis (PCA) was conducted based on the genotype data (Figure 6D) to reveal any potential population structure. The PCA revealed that the first principal component explained only 2.55% of the genetic variance, suggesting a low degree of dependence among markers. With the different categories of advancement labelled in different colors, the PCA showed a population structure with two separated clusters of the most and least developed material ("variety" and "cultivated", respectively). However, the mid-categories ("advanced" and "landrace") partially overlapped with the other categories. Interestingly, the accessions in the distinctly separated group of the least developed material originated exclusively from China (see Supplementary Figure 4 for a phylogenetic tree of the relatedness of the accessions).

3.6 GWAS identified markers related to ten different traits

A genome-wide association study (GWAS) was conducted to identify genetic loci associated with the 12 traits assessed in the diversity panel. Using the 6,606 SNPs mapped to the faba bean genome, 51 markers were identified that showed significant associations to 10 different traits: days to flowering, days to maturity, plant height, yield, number of seeds per plant, TGW, seed size, seed length, seed width, and weevil susceptibility (Table 2;

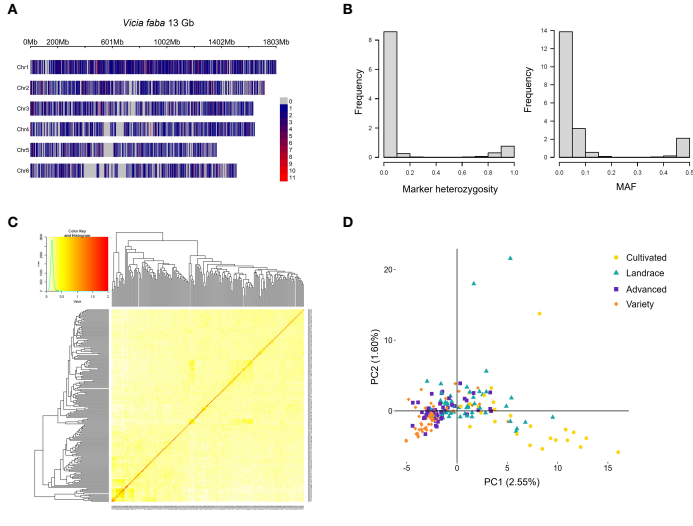


FIGURE 6
Genetics of the faba bean diversity panel. The markers obtained from DArTseq mapped to the 'Hedin/2' reference genome of *V. faba* with the color scale of the gradient indicating the marker density as number of SNPs in a 1Mbp window size **(A)**. Marker heterozygosity and minor allele frequency (MAF) **(B)**. Kinship matrix **(C)**. Principle component (PC) analysis plot based on genotype data and color-coded according to level of advancement **(D)**.

TABLE 2 Significant markers in V. faba that were associated with traits of interest identified through the GWAS, with a minor allele frequency (MAF) above 0.01.

Trait	Marker ID	Chr	Position (Mb)	Allele	p-value (log10)	MAF	Effect	Reference genome mapping Vfaba.Hedin2.R1	Sequence type	Gene annotation[a]	Homolog annotation[b]
DTF (days)	D10069	1S	248565604	C>T	1.00E05	0.058	2.24	1g038400	Exon	Dolichol kinase	Dolichol kinase EVAN isoform X1 [*Pisum sativum*]
	D18440	1L	857572779	T>C	4.50E13	0.138	3.12	–	Intergenic	–	–
	D13055	2	240493197	G>A	9.55E07	0.204	3.16	2g041560	Exon	Polyadenylate binding protein interacting protein	Zinc-finger BED domain-containing protein RICESLEEPER 1-like [*Arachis stenosperma*]
	D04593	4	1511466468	A>G	1.38E06	0.455	2.81	4g221040	Intron	Unknown protein	Unknown protein [*Pisum sativum*]
	D05439	5	723007937	G>T	2.42E06	0.094	1.55	5g104720	Intergenic[c]	Chromatin-remodeling ATPase, INO	Chromatin-remodeling ATPase, INO 80-like [*Pisum sativum*]
DTM (days)	D15934	1S	17355728	C>A	4.61E14	0.429	3.74	1g002800	Intron	Protein NRT/PTR family	Protein NRT1/ PTR FAMILY 6.1 [*Pisum sativum*]
	D08488	3	415004717	G>A	2.44E06	0.488	3.28	3g065720	Exon	Non-lysosomal glucosylceramidase	Non-lysosomal glucosylceramidase-like [*Trifolium medium*]
	D10812	3	521125812	C>G	9.58E06	0.071	1.43	3g084840	Intergenic[c]	phosphoglycerate mutase protein	phosphoglycerate mutase-like protein AT74 [*Pisum sativum*]
	D12639	4	527939461	C>T	2.16E05	0.452	1.78	4g093800	Intron	Translin-associated protein x homolog (TSNAX or TRAX)	translin-associated protein X-like [*Trifolium pratense*]
	D11891	4	808031589	G>A	1.93E08	0.429	2.20	–	Intergenic	–	
HEIGHT (cm)	D15935	1S	17355682	T>C	1.68E07	0.405	1.58	1g002800	Intron	Protein NRT/PTR family	Protein NRT1/ PTR FAMILY 6.1 [*Pisum sativum*]
	D13055	2	240493197	G>A	9.55E07	0.204	3.16	2g041560	Exon	Polyadenylate binding protein interacting protein	Zinc-finger BED domain-containing protein RICESLEEPER 1-like [*Arachis stenosperma*]
	D09581	3	40132566	A>T	1.84E06	0.479	8.91	3g013960	Intron	Casein kinase isoform	casein kinase 1-like protein 10 [*Vicia villosa*]
	D05633	4	5493793	C>G	9.10E06	0.469	8.84	4g001520	Intron	Alpha beta hydrolase	alpha/beta fold hydrolase [*Trifolium pratense*]
YIELD (g/plant)	D15935	1S	17355682	T>C	1.68E07	0.405	1.58	1g002800	Intron	Protein NRT/PTR family	Protein NRT1/ PTR FAMILY 6.1 [*Pisum sativum*]
	D01869	1S	539712640	A>G	6.84E06	0.065	1.23	1g080440	Exon	Phosphoenolpyruvate Carboxylase (PEPC)	phosphoenolpyruvate carboxylase [*Vicia villosa*]
	D18653	1L	1470877064	A>G	1.64E06	0.086	1.36		Intergenic	–	

(Continued)

10.3389/fpls.2024.1348014

TABLE 2 Continued

Trait	Marker ID	Chr	Position (Mb)	Allele	p-value (log10)	MAF	Effect	Reference genome mapping Vfaba.Hedin2.R1	Sequence type	Gene annotation[a]	Homolog annotation[b]
	D00223	3	483964717	A>G	1.18E07	0.120	1.06	3g076920	Exon	Protein Pelota homolog	CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase [*Trifolium repens*]
SEEDS (no/plant)	D08829	4	1393975921	G>T	3.53E06	0.094	3.28	4g204080	Exon	Fyve RHoGEF and PH domain-containing protein	FYVE zinc finger protein [*Medicago truncatula*] / vacuolar protein sorting-associated protein 27 [*Pisum sativum*]
TGW (g)	D08779	1S	716966083	G>A	3.35E11	0.063	1.53	1g105360	Exon	Salicylic acid glycosyl transferase, SGT homolog	protein SGT1 homolog [*Cicer arietinum*]
	D12875	1S	826078612	C>G	5.50E11	0.170	82.21	1g119840	Intron	Cytochrome c1-1 heme protein, mitochondrial	hypothetical protein TSUD_43780 [*Trifolium subterraneum*]
	D18440	1L	857572779	T>C	2.19E05	0.138	70.29	–	Intergenic	–	–
	D07779	1L	1336230743	A>T	2.39E06	0.085	86.00	1g407840	Exon	Methylthioribose phosphate isomerase	methylthioribose-1-phosphate isomerase isoform X1 [*Glycine soja*]
	D03153	2	581622870	A>G	1.08E07	0.074	1.12	2g099960	Exon	Gastric Triacylglycerol lipase	triacylglycerol lipase 2-like [*Vicia villosa*]
	D10357	2	1312826726	T>A	4.31E11	0.213	1.21	2g206480	Exon	Unknown protein	–
	D06119	2	1484092418	G>A	2.46E11	0.063	127.67	2g235160	Exon	Receptor Serine/Threonine Protein Kinase	probable serine/threonine-protein kinase PBL1 [*Vicia villosa*]
	D15752	4	1573846092	A>G	5.78E09	0.462	118.25	–	Intergenic	–	–
SIZE (mm^2)	D18111	1L	409615010	C>T	1.84E06	0.087	14.73	–	Intergenic	–	–
	D08779	1S	716966083	G>A	3.35E11	0.063	1.53	1g105360	Exon	Salicylic acid glycosyl transferase , SGT homolog	protein SGT1 homolog [*Cicer arietinum*]
	D12875	1S	826078612	C>G	5.50E11	0.170	82.21	1g119840	Intron	Cytochrome c1-1 heme protein, mitochondrial	hypothetical protein TSUD_43780 [*Trifolium subterraneum*]
	D07779	1L	1336230743	A>T	2.39E06	0.085	86.00	1g407840	Exon	Methylthioribose phosphate isomerase	methylthioribose-1-phosphate isomerase isoform X1 [*Glycine soja*]
	D03153	2	581622870	A>G	1.08E07	0.074	1.12	2g099960	Exon	Gastric Triacylglycerol lipase	triacylglycerol lipase 2-like [*Vicia villosa*]
	D10357	2	1312826726	T>A	4.31E11	0.213	1.21	2g206480	Exon	Unknown protein	–
	D06119	2	1484092418	G>A	2.88E17	0.063	28.44	2g235160	Exon	Receptor Serine/Threonine Protein Kinase	probable serine/threonine-protein kinase PBL1 [*Vicia villosa*]

(Continued)

10.3389/fpls.2024.1348014

TABLE 2 Continued

Trait	Marker ID	Chr	Position (Mb)	Allele	p-value (log10)	MAF	Effect	Reference genome mapping Vfaba.Hedin2.R1	Sequence type	Gene annotation[a]	Homolog annotation[b]
	D07466	3	401377640	C>A	3.23E06	0.491	46.80	–	Intergenic	–	–
	D15212	4	704354594	G>A	6.88E07	0.056	15.20	4g106080	Intron	Kinasine protein kin	kinesin-like protein KIN-14I [*Lotus japonicus*]
SEED LENGTH (mm)	D08779	1S	716966083	G>A	3.35E11	0.063	1.53	1g105360	Exon	Salicylic acid glycosyl transferase, SGT homolog	protein SGT1 homolog [*Cicer arietinum*]
	D08133	2	496103881	A>G	2.90E05	0.481	2.63	2g084080	Intron	2,3- bisphosphoglycerate dependent phosphoglycerate mutase (dPGAM)	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase [*Glycine max*]
	D03153	2	581622870	A>G	1.08E07	0.074	1.12	2g099960	Exon	Gastric Triacylglycerol lipase	triacylglycerol lipase 2-like [*Vicia villosa*]
	D10357	2	1312826726	T>A	4.31E11	0.213	1.21	2g206480	Exon	Unknown protein	–
	D07408	3	916408049	G>A	8.56E06	0.085	1.06	–	Intergenic	–	–
	D05388	3	1634506611	C>A	4.14E07	0.060	1.65	3g239080	Exon	GDSL esterase lipase	GDSL esterase/lipase At1g23500-like [*Vicia villosa*]
	D04763	5	404936671	T>C	3.82E06	0.059	1.43	5g063000	Intron	Human glycosyltransferase domain-containing protein	polygalacturonate 4-alpha-galacturonosyltransferase-like [*Pisum sativum*]
	D03671	5	992670147	T>C	1.52E06	0.075	1.12	5g142480	Intron	PWWP domain isoform	hypothetical protein KIW84_031546 [*Pisum sativum*]
SEED WIDTH (mm)	D08779	1S	716966083	G>A	3.35E11	0.063	1.53	1g105360	Exon	Salicylic acid glycosyl transferase , SGT homolog	protein SGT1 homolog [*Cicer arietinum*]
	D03153	2	581622870	A>G	1.08E07	0.074	1.12	2g099960	Exon	Gastric Triacylglycerol lipase	triacylglycerol lipase 2-like [*Vicia villosa*]
	D10357	2	1312826726	T>A	5.25E09	0.213	0.84	2g206480	Exon	Unknown protein	–
	D08952	5	1106041368	G>A	1.07E05	0.074	0.83	5g157800	Intron	Rhomboid protein	rhomboid-like protein 15 [*Vicia villosa*]
WEEVIL (% infested seeds)	D06119	2	1484092418	G>A	1.56E07	0.063	8.67	2g235160	Exon	Receptor Serine/Threonine Protein Kinase	probable serine/threonine-protein kinase PBL1 [*Vicia villosa*]
	D05758	3	1356117551	C>T	9.95E06	0.118	5.99	3g196560	Exon	Protein IQ-domain	protein IQ-DOMAIN 21-like isoform X2 [*Vicia villosa*]
	D00463	6	924462796	A>T	1.89E08	0.053	10.42	6g111840	Exon	Mediator of RNA polymerase II transcription subunit	mediator of RNA polymerase II transcription subunit 16 isoform X3 [*Vicia villosa*]

[a]Annotation based on reference genome annotation.
[b]Annotation based on highest sequence similarity of legume species using blastX at NCBI (BLAST).
[c]Intergenic, but located less than 300bp upstream of a start codon.

10.3389/fpls.2024.1348014

Figure 7; Supplementary Figure 5 for starch and protein content). The quantile-quantile (QQ) plots of the GWAS with the BLINK model showed a strong association between the predicted and expected distributions of the p-values suggesting that the models appropriately accounted for the population stratification. There were no overlaps between markers for days to flowering and days to maturity. However, two markers associated with days to flowering (D13055 and D18440) were also associated with plant height and seed weight, respectively. Not surprisingly, several markers associated with different seed dimension traits showed overlaps. Of the three markers associated with weevil damage, one (D06119) was also associated with seed weight and size.

3.7 Genomic localization of markers and candidate genes

The majority of the markers, 40 out of 51, were localized to predicted gene coding regions of the reference genome (Jayakodi et al., 2023). Annotations for these genes were derived from the reference genome and, for confirmation purposes, from genes showing the highest sequence similarity in legume species (Table 2). In most cases, these annotations were consistent with each other. The candidate genes encode proteins with functions in central biosynthetic pathways, DNA/RNA metabolism, and transcriptional regulation or as hormone and metabolite transporters.

4 Discussion

4.1 Importance of diversity panels in pre-breeding

In this work, we assembled a diversity panel of 220 faba bean accessions including a wide range of geographical origins and breeding advancements. This panel was characterized through genotyping and field phenotyping under Nordic climate conditions, in this case a warm-summer humid continental climate (Arnfield, 2023). We genotyped our diversity panel using the genotyping-by-sequencing (GBS) approach DArTseq to identify single nucleotide polymorphism (SNP) markers (Baloch et al., 2017). Based on this genotype data it became evident that the accessions exhibited a very low degree of relatedness as supported by the kinship matrix and the principal component analysis (PCA) indicated a low dependence among markers. The comprehensive phenotypic characterization of this diversity panel through field

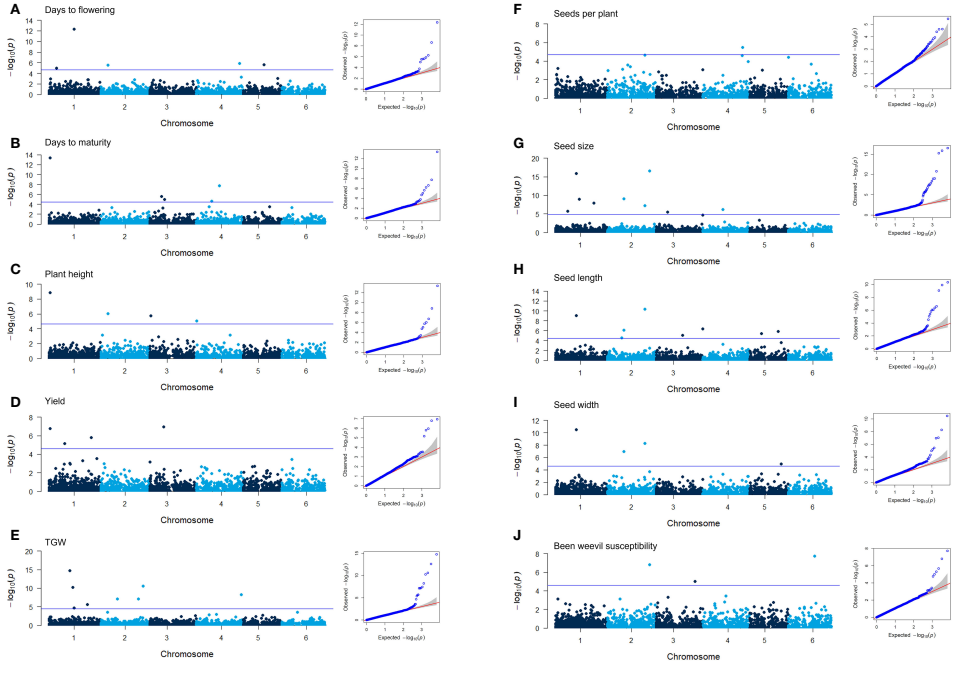


FIGURE 7
Manhattan plots of the GWAS results for 10 of the traits characterized in the diversity panel of *V. faba* and the corresponding QQ plots **(A–J)**. The blue line in the Manhattan plot signifies the false discovery rate (FDR) threshold. QQ plots showing the distribution of p-values of the markers tested for trait association in the GWAS using the BLINK model plotted against the expected distribution of p-values.

trials conducted over two consecutive years revealed substantial variation in key agronomic and seed quality traits (days to flowering, days to maturity, plant height, bean weevil susceptibility, seed yield, thousand grain weight, number of seeds per plant, seed size, seed width, seed length, protein content, and starch content). The large variation in seed size of *Vicia faba* is well known and has historically led to its categorization into subspecies major (large-seeded), equine (medium-seeded), minor (small-seeded), and paucijuga (small-seeded) types. However, due to the phenotypic continuity rather than distinctness, and their reproductive compatibility, these types are today regarded botanical groups of the same species (Maalouf et al., 2018; Jayakodi et al., 2023). Several of the characterized traits showed high broad-sense heritability values, indicating a strong additive genetic component. The heritability of seed size traits in our study was very high and similar to those seen in previous work with values close to 1.00 (Skovbjerg et al., 2023). Our results from characterization of a diversity panel of faba bean present a valuable resource for identifying germplasm with desirable traits for breeding purposes. It can be noted that extending the characterization of this diversity panel across multiple years and locations would be beneficial to confirm the genetic impact on specific traits.

The phenotype and genotype data were further used in a genome-wide association study (GWAS) to identify marker-trait associations. Our SNP marker data comprised a high rate of rare alleles (Figure 6B). It should be noted that the choice of MAF threshold affects the power of GWAS and is commonly set to 0.05 (Alqudah et al., 2020). However, to capture rare alleles in diverse panels which can be of interest as targets for further genetic and biological studies, the MAF threshold in the GWAS can be lowered to not exclude any potential markers of interest (Stanton-Geddes et al., 2013; Arkwazee et al., 2022; Kim et al., 2022). In the current study, the MAF was therefore set to >0.01, however, it is noteworthy, that the markers identified as significantly associated with traits all had a MAF exceeding 0.05. While rare alleles could lead to spurious associations in GWAS as they are present in very few individuals, it is worth noting that some rare alleles may control important traits. Therefore, rare alleles in GWAS require careful attention (Gibson, 2012). For example, rare alleles were found to be associated with important agronomic traits in wheat (Jaiswal et al., 2016), and grain size and yield in rice (Hu et al., 2015). It is important that the associated markers with rare alleles identified in this study, are validated in different genetic backgrounds or larger populations before being used in marker-assisted selection to improve agronomic and yield traits in faba bean. Overall, a less diverse panel (in terms of level of advancement, germplasm origin, and relatedness) or a larger number of accessions would have been beneficial to capture a more extensive set of associated markers, since these factors impair the genetic analysis of the traits of interest (Torkamaneh and Belzile, 2022). However, the aim of this study was to identify novel loci of interest for trait screening instead of the development of markers for direct implementation in genomic breeding of advanced lines.

4.2 Phenotypic breeding advancement and target traits for Nordic regions

For the Nordic region which is characterized by a short growing season, the relatively long growth period of faba bean as compared to other spring-sown crops emphasizes the significance of prioritizing earliness traits as a key breeding target (Stoddard and Hämäläinen, 2011). Nevertheless, we showed that previous breeding advances in the faba bean germplasm have indeed resulted in notable improvements such as higher yields and a greater number of seeds per plant, but also a later onset of flowering and maturity. It is expected that the modern varieties were higher yielding, considering that most of the modern varieties in our diversity panel were developed for central or southern Europe, a climate region fairly similar to the latitude of our field trials, as compared to the wider geographical origin of the less improved material. Furthermore, given that the growing season is relatively long in southern and central Europe and that earliness traits have thus not been prioritized in breeding programs of faba bean in this region, it is not surprising that modern varieties in our diversity panel showed a later onset of flowering and maturation.

Our phenotypic correlation studies based on all accessions in the diversity panel showed a significant positive relationship between late maturity and higher yields, and a higher number of seeds per plant. However, our observation of broad variation in earliness traits, with more than two weeks difference in days to flowering as well as in days to maturity, signifies that relevant germplasm is available for breeding aimed at earlier varieties. The accessions showing the earliest flowering originated from diverse geographical regions including the Nordic latitudes, and represented several landraces (for example Anuksen Kanta, Brottby, Grebo, and Gubbestad) but also vegetable varieties available for hobby growers (such as Crimson flowered, Hangdown, and Robin Hood). Our pairwise correlation analysis revealed a strong correlation between early flowering and large seed size, of which the latter is usually not a desirable trait for cultivation on larger scale, due to their incompatibility with sowing and harvest machines and lower seed amplification factor. However, among the accessions in our diversity panel it is possible to identify accessions combining early flowering (less than 57 days to flowering), reasonable seed size (thousand-grain weight (TGW) between 400-700 g) and high yield (>10 g/plant) such as Cervci, Talia, FAB5138, FAB6776, Brottby, Jygeva, Gubbestad, and Habas de Beck).

4.3 Food quality aspects of faba bean seeds are important breeding targets

The increased application of faba bean for human consumption is relying on its nutritional composition (Martineau-Côté et al., 2022). In our work, post-harvest seed quality aspects were characterized alongside the agronomic traits, shedding light on the variability of traits such as protein and starch content of, 25-38% and 21-38% by seed dry weight, respectively. These ranges are in line with reports from

other faba bean studies (Martineau-Côté et al., 2022; Zhao et al., 2023). Our findings further indicated that there was no observed yield penalty for increased seed protein content, under our experimental field conditions. This stands in contrast to what is well-known for wheat, for example, where seed protein content is negatively correlated with yield (Oury and Godin, 2007; Laidig et al., 2017). On the other hand, in pea, both negative and positive correlations between seed protein content and yield have been reported (Daba and Morris, 2022). Our results showed a negative correlation between protein and starch content in faba bean, which has also been observed in pea (Daba and Morris, 2022), implying that breeding for increased protein content potentially leads to a reduction of starch. Furthermore, our results showed that TGW was negatively correlated to both protein and starch content, indicating that smaller seeds have a higher protein and starch content.

In recent times broad bean weevil (*Bruchus rufimanus*) infestation of faba bean seeds has emerged as a severe pest (Huber et al., 2023). Despite the rather strict quality standards in place with a maximum acceptance of 3% infested seeds for human consumption and less than 10% for animal feed, broad bean weevil resistance remains elusive, and the understanding of chemical attractors associated with the pest is still limited (Dell'Aglio and Tayeh, 2023). The overall weevil infestation was high in our diversity panel. However, our data uncovered a gradient among accessions, suggesting varying levels of susceptibility to bean weevil damage. This observation aligns with prior studies that have highlighted varietal differences in susceptibility to bean weevil damage (Carrillo-Perdomo et al., 2019; Segers et al., 2022; Dell'Aglio and Tayeh, 2023). Based on the results in our study and in previously published data (Dell'Aglio and Tayeh, 2023), selecting for traits such as low plant height, late flowering and maturation and small seed size could be a possible approach to mitigate weevil infestation, since they were all negatively correlated to weevil damage of seeds. However, it should be considered that the early flowering accessions can act as catch crops. In fact, Dell'Aglio and Tayeh (2023) suggested that earlier flowering plants may offer pods earlier while larger seeds provide more food for the larvae, potentially contributing to a higher susceptibility to weevil infestation. Other studies have shown that bean weevil damage reduces yields in faba bean (Carrillo-Perdomo et al., 2019; Segers et al., 2022) but our study did not show this correlation. It should be noted that our plant material exhibited a high diversity and multiple factors with small effects probably contribute to the varying degree of bean weevil susceptibility. Further studies are necessary to develop effective strategies for managing and overcoming this pest in faba bean production.

4.4 Novel markers for ten different traits identified in candidate genes

Several reports on different crops including legumes have shown the potential of using DArTseq to determine population structures and associating genetic markers with different traits (Akbari et al., 2006; Raman et al., 2014; Aznar-Fernández et al., 2020; Alemu et al., 2022). By using DArTseq in this population we could identify 6,606 SNP markers evenly distributed in the faba bean genome, and 51 of those markers were identified to be associated with ten of the characterized traits through our GWAS. Among these traits, seed size characters and TGW were predominantly represented in the list of markers, along with yield, number of seeds per plant, plant height, days to flowering, days to maturation, and bean weevil damage.

Three of the markers associated with days to flowering were localized to predicted genes with homologs in other plant species encoding proteins that regulate flowering; namely dolichol kinase, polyadenylate binding protein interacting protein, and chromatin-remodeling ATPase INO (Zhang Y. et al., 2015; Cho et al., 2017; Wang et al., 2019). Those markers have not been reported in previous marker-trait studies on faba bean targeting flowering traits (Aguilar-Benitez et al., 2021; Skovbjerg et al., 2023). The chromatin-remodeling ATPase INO80 has a prominent role in the regulation of the key flowering repressor *FLOWERING LOCUS C*, and Arabidopsis *ino80* mutants show a delayed onset of flowering (Zhang C. et al., 2015). Interestingly, the marker identified in our study was localized not within but 209 bp upstream of the predicted gene, with a high probability of being part of its regulatory region. Days to flowering is physiologically connected to days to maturation, and our phenotype data showed a strong positive correlation between them. In our study, markers associated with days to maturation were localized to predicted genes for which homologs in other plant species have functions in plant hormone transport, sphingolipid metabolism, glycolysis, and RNA/DNA metabolism (Zhao and Assmann, 2011; Chiba et al., 2015; Gupta et al., 2017; Dai et al., 2020).

One of the four markers associated with plant height was also associated to yield, and only 46 bp away from a marker associated with days to maturation, all localized to the same predicted gene involved in hormone transport. Another marker associated with plant height was also associated with days to flowering. The genetic markers linked to plant height as identified by Skovbjerg et al. (2023) were not identical to the markers in our dataset. Two of the markers associated to seed yield in our study were localized to predicted genes with homologs in other plant species being involved in glycolysis and lipid metabolism. One of them encodes a phosphoenolpyruvate carboxylase that has an important role in carbon and nitrogen metabolism, and has recently been linked to seed yield in Arabidopsis (Shi et al., 2015; Feria et al., 2022).

Of the six markers associated with TGW, one was localized to a predicted gene encoding a salicylic acid glycosyl transferase. Salicylic acid is a phytohormone that is usually connected to functions in biotic and abiotic stress response but has, interestingly, also been shown to regulate the number of flowers and pods in chickpea (Li et al., 2022). Another marker was localized to a predicted gene for which the homolog in Arabidopsis encodes methylthioribose phosphate isomerase shown to be involved in promoting flower and seed development (Zierer et al., 2016). As expected, a large number of the markers associated with TGW were also associated with the seed dimension traits area, length, and width. Skovbjerg et al. (2023) previously identified similar marker overlaps for various seed dimension parameters in faba beans. However, none of the genes mapped by their markers were

present in our dataset. The three markers found by Zhao et al. (2023) did not match our markers for TGW and seed size features. Historically, the selection of faba bean likely relied solely on visual traits of the seeds' morphology. Thus, the discovery of genetic markers related to seed characteristics might be promising for further breeding endeavors.

One of the three markers associated with bean weevil damage was also associated with seed size. In fact, a correlation between seed size and bean weevil susceptibility was observed in our phenotype data where larger seeds had a higher ratio of bean weevil damage. The second marker was localized to a predicted gene for which an Arabidopsis homolog encodes an IQ domain protein with a function in calcium ion signaling (Fischer et al., 2013), which is important in biotic stress response (Aldon et al., 2018). The third marker was localized to a gene homologous to a subunit of the mediator complex of RNA polymerase II which in Arabidopsis is involved in transcriptional responses to cell wall defects (Buendía-Monreal and Gillmor, 2016). These functions can potentially be involved in the response to bean weevil attacks.

Despite the large variation observed in seed protein and starch content in our diversity panel, the GWAS did not reveal any markers associated with these traits. Recently, Zhao et al. (2023) revealed 22 markers associated with seed quality traits, including seed protein, starch, and lipid content. It is noteworthy that those markers were identified using a genotyping platform with a 130 K SNP chip based on RNA data from flower and leaf tissue but not from seed tissue (Wang et al., 2021).

The absence of overlap between our associated markers with earlier reported markers in faba bean is not entirely surprising, when taking into consideration the species vast genome and the relatively few available studies on faba bean so far. Further, our diversity panel presents a heterogenous population with a broad genetic base compared to the more narrow genetic bases of other studies.

4.5 Challenges with genotyping of faba bean being a partly outcrossing species

Genotyping by sequencing (GBS) approaches has the potential to identify novel SNP at a high density which can improve genomic analyses such as GWAS, genomic prediction and QTL mapping (Meuwissen and Goddard, 2010; Druet et al., 2014; Alipour et al., 2019). However, many GBS approaches suffer from an abundance of missing data due to low sequencing depth which can significantly lower the number of usable SNPs. DArTseq relies on a combination of restriction enzymes for the digestion of the genome prior to the sequencing of the resulting segments (https://www.diversityarrays.com/). One explanation for the large proportion of the missing data, of the original 19,770 SNP markers, could be that no SNP call could be produced for segments in the accessions where the upstream restriction enzyme site was not present and therefore not sequenced. Through the DArTseq method it was possible to identify a large amount of polymorphisms in the genome that could potentially be used for development of marker chips, however for the genotyping of multiple individuals in a large diverse population, sequencing methods based on digestion with restriction enzymes might not be preferable.

Increasing sequencing depth has the potential to lower the amount of missing data through higher repeatability in marker sequence reads and coverage of the genome, simply through the increase of reads per sample but at higher costs. Instead, improvement of the data through imputation of the missing values could increase the usability of the genetic data. Using close neighbors to impute missing genotyping data from GBS platforms is an efficient strategy which increases the accuracy of downstream genetic analyses (Gamal El-Dien et al., 2015). In this study the proportion of missing data was 29.6% after filtration steps, which was imputed using the five closest neighbors. While our study successfully identified 51 novel markers associated to ten different traits, it is important to acknowledge the potential impact of the genotyping challenges and the approach of genotyping pooled individuals. The majority of associated markers were located on chromosome 1, the largest among the six chromosomes. Interestingly, no markers were identified on chromosome 6, despite the generally even distribution of the 6,606 SNP markers across the faba bean genome. The relatively low marker coverage on individual chromosomes, considering the vast 13 Gb genome of faba bean, may account for this phenomenon. Increasing the sequencing depth could potentially have resulted in a higher density of informative markers, enhancing the dataset's comprehensiveness and robustness. This, in turn, might have facilitated the identification of more markers associated with traits of interest.

As mentioned earlier, the relatively high proportion of missing data could partly be addressed through imputation. However, the partly outcrossing nature of faba bean, leading to increased heterozygosity within accessions, raises concerns about accurately determining genotypes. To account for this, genotyping of each accession in the panel was performed on five pooled individuals aiming to determine the genotype of that 'population' rather than of individuals. In general, to capture a representative sample of the genetic variation within a population, genotyping a larger number of individuals for each accession would have been beneficial. Further, potential misclassification of genotypes as homozygotes or heterozygotes for the SNPs, possibly caused by handling SNP data from pooled individuals (in which the individuals' genotypes cannot be properly identified), could influence the interpretation of marker-trait associations. For example, if minor alleles are distributed evenly within, rather than among accessions, this might wrongly infer homozygotes at many loci. However, rare alleles present at a low frequency within an accession might only contribute to a small proportion of the genetic effect on the resulting phenotype. Therefore, while the novel markers from our study show promise in being associated with target traits and to genes with predicted functions in key developmental processes in plants, further evaluation is needed to assess their reliability and practical utility in breeding programs.

5 Conclusions

Recent breeding efforts on faba bean are scanty in comparison to other legumes, such as soybean or pea (Rubiales et al., 2021). However, the recently published reference genome of faba bean (Jayakodi et al., 2023) has opened new possibilities for exploring the genetics of faba bean with significantly higher resolution and efficiency and how this can be used in future plant breeding. To establish a sustainable supply of locally produced green proteins, concerted breeding efforts are crucial for faba bean targeting different geographical regions and leveraging genetically informed breeding techniques. To address these challenges, plant breeders can harness the vast genetic diversity available in faba bean germplasm collections held in genebanks (Duc et al., 2010; Rubiales et al., 2021). This necessitates an investment in pre-breeding, which involves the thorough characterization of germplasm through both phenotypic and genotypic assessments, as exemplified by the findings presented in this study. The results from this study contribute to the growing pool of phenotypic and genotypic data on faba bean, which provides a valuable resource for developing efficient breeding strategies for faba bean.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

HO: Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. JÅ: Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. AlC: Conceptualization, Validation, Writing – review & editing. DB: Conceptualization, Writing – review & editing. CH: Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. AaC: Conceptualization, Investigation, Methodology, Supervision, Validation, Writing – review & editing. ÅG: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was made possible by funding from SLU Grogrund 'Future faba beans for food and feed' (#ltv.2018.1.1.1.1067), SLU LTV faculty for supporting growth facilities, the Swedish Infrastructure for Ecosystem Science (SITES) Lönnstorp Research Station at SLU (Swedish Research Council 2017-00635), and The Royal Physiographic Society in Lund, Sweden, for analytical equipment.

Acknowledgments

We would like to thank research assistants Kaye Bateman, Anna Nieto Esteve, and Jenny Östberg for nursing plants, seed logistics, and phenotypic characterization in field and lab.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1348014/full#supplementary-material

References

Abou-Khater, L., Maalouf, F., Jighly, A., Alsamman, A. M., Rubiales, D., Rispail, N., et al. (2022). Genomic regions associated with herbicide tolerance in a worldwide faba bean (*Vicia faba* L.) collection. *Sci. Rep.* 12, 158. doi: 10.1038/s41598-021-03861-0

Aguilar-Benitez, D., Casimiro-Soriguer, I., Maalouf, F., and Torres, A. M. (2021). Linkage mapping and QTL analysis of flowering time in faba bean. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-021-92680-4

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., et al. (2006). Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.* 113, 1409–1420. doi: 10.1007/s00122-006-0365-4

Aldon, D., Mbengue, M., Mazars, C., and Galaud, J. P. (2018). Calcium signalling in plant biotic interactions. *Int. J. Mol. Sci.* 19, 665. doi: 10.3390/ijms19030665

Alemu, A., Brantestam, A. K., and Chawade, A. (2022). Unraveling the genetic basis of key agronomic traits of wrinkled vining pea (*Pisum sativum* L.) for sustainable production. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.844450

Alipour, H., Bai, G., Zhang, G., Bihamta, M. R., Mohammadi, V., and Peyghambari, S. A. (2019). Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. *PloS One* 14, e0208614. doi: 10.1371/journal.pone.0208614

Alqudah, A. M., Sallam, A., Baenziger, P. S., and Börner, A. (2020). GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley - A Review. *J. Adv. Res.* 22, 119–135. doi: 10.1016/j.jare.2019.10.013

Alvarado, G., Rodriguez, F. M., Pacheco, A., Burgueño, J., Crossa, J., Vargas, M., et al. (2020). META-R: A software to analyze data from multi-environment plant breeding trials. *Crop J.* 8, 745–756. doi: 10.1016/j.cj.2020.03.010

Arkwazee, H. A., Wallace, L. T., Hart, J. P., Griffiths, P. D., and Myers, J. R. (2022). Genome-wide association study (GWAS) of white mold resistance in snap bean. *Genes* 13. doi: 10.3390/genes13122297

Arnfield, A. J. (2023)Köppen climate classification. Available online at: https://www.britannica.com/science/Koppen-climate-classification (Accessed 18 January 2024).

Aznar-Fernández, T., Barilli, E., Cobos, M. J., Kilian, A., Carling, J., and Rubiales, D. (2020). Identification of quantitative trait loci (QTL) controlling resistance to pea weevil (*Bruchus pisorum*) in a high-density integrated DArTseq SNP-based genetic map of pea. *Sci. Rep.* 10, 1–12. doi: 10.1038/s41598-019-56987-7

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Çiftçi, V., Sáenz De Miera, L. E., Aasim, M., et al. (2017). A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. *PloS One* 12, 1–18. doi: 10.1371/journal.pone.0167821

Björnsdotter, E., Nadzieja, M., Chang, W., Escobar-Herrera, L., Mancinotti, D., Angra, D., et al. (2021). VC1 catalyses a key step in the biosynthesis of vicine in faba bean. *Nat. Plants* 7, 923–931. doi: 10.1038/s41477-021-00950-w

BLAST *BLAST: Basic Local Alignment Search Tool*. Available online at: https://blast.ncbi.nlm.nih.gov/Blast.cgi (Accessed 2023-11-09).

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Buendia-Monreal, M., and Gillmor, C. S. (2016). Mediator: A key regulator of plant development. *Dev. Biol.* 419, 7–18. doi: 10.1016/j.ydbio.2016.06.009

Carrillo-Perdomo, E., Raffiot, B., Ollivier, D., Deulvot, C., Magnin-Robert, J. B., Tayeh, N., et al. (2019). Identification of novel sources of resistance to seed weevils (*Bruchus* spp.) in a faba bean germplasm collection. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01914

Carrillo-Perdomo, E., Vidal, A., Kreplak, J., Duborjal, H., Leveugle, M., Duarte, J., et al. (2020). Development of new genetic resources for faba bean (*Vicia faba* L.) breeding through the discovery of gene-based SNP markers and the construction of a high-density consensus map. *Sci. Rep.* 10, 1–14. doi: 10.1038/s41598-020-63664-7

Chiba, Y., Shimizu, T., Miyakawa, S., Kanno, Y., Koshiba, T., Kamiya, Y., et al. (2015). Identification of *Arabidopsis thaliana* NRT1/PTR FAMILY (NPF) proteins capable of transporting plant hormones. *J. Plant Res.* 128, 679–686. doi: 10.1007/s10265-015-0710-2

Cho, Y., Yu, C.-Y., Nakamura, Y., Kanehara, K., and Murphy, A. (2017). Arabidopsis dolichol kinase AtDOK1 is involved in flowering time control. *J. Exp. Bot.* 68, 3243–3252. doi: 10.1093/jxb/erx095

CPVO *CPVO*. Available online at: https://online.plantvarieties.eu/publicSearch (Accessed 2023-11-09).

Daba, S. D., and Morris, C. F. (2022). Pea proteins: Variation, composition, genetics, and functional properties. *Cereal Chem.* 99, 8–20. doi: 10.1002/cche.10439

Dai, G.-Y., Yin, J., Li, K.-E., Chen, D.-K., Liu, Z., Bi, F.-C., et al. (2020). The Arabidopsis AtGCD3 protein is a glucosylceramidase that preferentially hydrolyzes long-acyl-chain glucosylceramides. *J. Biol. Chem.* 295, 717–728. doi: 10.1016/S0021-9258(17)49930-3

Dell'Aglio, D. D., and Tayeh, N. (2023). Responsiveness of the broad bean weevil, *Bruchus rufimanus*, to *Vicia faba* genotypes. *Entomol. Experimentalis Applicata*. 171:312–322 doi: 10.1111/EEA.13277

Druet, T., Macleod, I. M., and Hayes, B. J. (2014). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Hered. (Edinb.)* 112, 39–47. doi: 10.1038/hdy.2013.13

Duc, G., Bao, S., Baum, M., Redden, B., Sadiki, M., Suso, M. J., et al. (2010). Diversity maintenance and use of *Vicia faba* L. genetic resources. *Field Crops Res.* 115, 270–278. doi: 10.1016/j.fcr.2008.10.003

Ellwood, S. R., Phan, H. T. T., Jordan, M., Hane, J., Torres, A. M., Avila, C. M., et al. (2008). Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with Lens culinaris. *BMC Genomics* 9, 380. doi: 10.1186/1471-2164-9-380

FAOSTAT (2021) *License: CC BY-NC-SA 3.0 IGO*. Available online at: https://www.fao.org/faostat/en/#compare (Accessed 2023-04-26).

Feria, A. B., Ruíz-Ballesta, I., Baena, G., Ruíz-López, N., Echevarría, C., and Vidal, J. (2022). Phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxylase kinase isoenzymes play an important role in the filling and quality of *Arabidopsis thaliana* seed. *Plant Physiol. Biochem.* 190, 70–80. doi: 10.1016/j.plaphy.2022.08.012

Fischer, C., Kugler, A., Hoth, S., and Dietrich, P. (2013). An IQ domain mediates the interaction with calmodulin in a plant cyclic nucleotide-gated channel. *Plant Cell Physiol.* 54, 573–584. doi: 10.1093/pcp/pct021

Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Chen, C., Porth, I., and El-Kassaby, Y. A. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genomics* 16, 1–16. doi: 10.1186/s12864-015-1597-y

Gibson, G. (2012). Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* 13, 135–145. doi: 10.1038/nrg3118

Gupta, A., Nair, A., Ballal, A., and Chittela, R. K. (2017). C-terminal residues of rice translin are essential for octamer formation and nucleic acid binding. *Plant Physiol. Biochem.* 118, 600–608. doi: 10.1016/j.plaphy.2017.08.004

Hu, J., Wang, Y., Fang, Y., Zeng, L., Xu, J., Yu, H., et al. (2015). A rare allele of GS2 enhances gain size and grain yield in rice. *Mol. Plant* 8, 1455–1465. doi: 10.1016/j.molp.2015.07.002

Huber, J., Chaluppa, N., Voit, B., Steinkellner, S., and Killermann, B. (2023). Damage potential of the broad bean beetle (Bruchus rufimanus Boh.) on seed quality and yield of faba beans (*Vicia faba* L.). *Crop Prot.* 168, 106227. doi: 10.1016/j.cropro.2023.106227

Jaiswal, V., Gahlaaut, V., Meher, P. K., Mir, R. R., Jaiswal, J. P., Rao, A. R., et al. (2016). Genome wide single locus single trait, multi-locus and multitrait association mapping for some important agonomic traits in common wheat (*T. aestivum* L.). *PloS One* 11, e0159343. doi: 10.1371/journal.pone.0159343

Jayakodi, M., Golicz, A. A., Kreplak, J., Fechete, L. I., Angra, D., Bednář, P., et al. (2023). The giant diploid faba genome unlocks variation in a global protein crop. *Nature* 615, 652–659. doi: 10.1038/s41586-023-05791-5

Karaköy, T., Toklu, F., Karagöl, E. T., Uncuer, D., Çilesiz, Y., Ali, A., et al. (2023). Genome-wide association studies revealed DArTseq loci associated with agronomic traits in Turkish faba bean germplasm. *Genet. Resour. Crop Evol.* 71, 181–198. doi: 10.1007/s10722-023-01615-7

Khazaei, H., O'Sullivan, D. M., Stoddard, F. L., Adhikari, K. N., Paull, J. G., Schulman, A. H., et al. (2021). Recent advances in faba bean genetic and genomic tools for crop improvement. *Legume Sci.* 3(3), e75. doi: 10.1002/leg3.75

Kim, J. M., Lyu, J., Kim, D. G., Hung, N. N., Seo, J. S., Ahn, J. W., et al. (2022). Genome wide association study to detect genetic regions related to isoflavone content in a mutant soybean population derived from radiation breeding. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.968466

Laidig, F., Piepho, H.-P., Rentel, D., Drobek, T., Meyer, U., and Huesken, A. (2017). Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983-2014. *Theor. Appl. Genet.* 130, 223–245. doi: 10.1007/s00122-016-2810-3

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, M. W., He, Y. H., Liu, R., Li, G., Wang, D., Ji, Y. S., et al. (2023). Construction of SNP genetic maps based on targeted next-generation sequencing and QTL mapping of vital agronomic traits in faba bean (*Vicia faba* L.). *J. Integr. Agric.* 22, 2648–2659. doi: 10.1016/j.jia.2023.01.003

Li, A., Sun, X., and Liu, L. (2022). Action of salicylic acid on plant growth. *Front. Plant Sci.* 13, 878076. doi: 10.3389/fpls.2022.878076

Maalouf, F., Abou-Khater, L., Babiker, Z., Jighly, A., Alsamman, A. M., Hu, J., et al. (2022). Genetic dissection of heat stress tolerance in faba bean (*Vicia faba* L.) using GWAS. *Plants* 11, 1–17. doi: 10.3390/plants11091108

Maalouf, F., Hu, J., O'Sullivan, D. M., Zong, X., Hamwieh, A., Kumar, S., et al. (2018). "Breeding and genomics status in faba bean (*Vicia faba*)," in *Plant Breeding*. Ed. C. Ojiewo (John Wiley & Sons, Ltd). 138, 465–473 doi: 10.1111/pbr.12644

Martineau-Côté, D., Achouri, A., Karboune, S., and L'Hocine, L. (2022). Faba bean: an untapped source of quality plant proteins and bioactives. *Nutrients* 14, 1541. doi: 10.3390/nu14081541

Meuwissen, T., and Goddard, M. (2010). Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 185, 623–631. doi: 10.1534/genetics.110.116590

Mohsin, A. (2021) *LD decay plot from TASSEL LDoutput.R (Version 1)*. Available online at: https://github.com/mohsinali1990/My_scripts/blob/c3ef6d56cbeb40d9c96eb79e869a645d6d8b2052/LD%20decay%20Plot%20from%20TASSEL%20LDoutput.R.

Oury, F. X., and Godin, C. (2007). Yield and grain protein concentration in bread wheat: How to use the negative relationship between the two characters to identify favourable genotypes? *Euphytica* 157, 45–57. doi: 10.1007/s10681-007-9395-5

Raman, H., Raman, R., Kilian, A., Detering, F., and Carling, J. (2014). Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PloS One* 9, 101673. doi: 10.1371/journal.pone.0101673

Remington, D. L., Thornsberry, M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., et al. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci.* 98, 11479–11484. doi: 10.1073/pnas.201394398

Rubiales, D., Annicchiarico, P., Vaz Patto, M. C., and Julier, B. (2021). Legume breeding for the agroecological transition of global agri-food systems: A European perspective. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.782574

Segers, A., Dumoulin, L., Megido, R. C., Jacquet, N., Cartrysse, C., Kamba, P. M., et al. (2022). Varietal and environmental effects on the production of faba bean (*Vicia faba* L.) seeds for the food industry by confrontation of agricultural and nutritional traits with resistance against Bruchus spp. (Coleoptera: Chrysomelidae, BruChinae). *Agric. Ecosyst. Environ.* 327, 107831. doi: 10.1016/j.agee.2021.107831

Shi, J., Yi, K., Liu, Y., Xie, L., Zhou, Z., Chen, Y., et al. (2015). Phosphoenolpyruvate carboxylase in Arabidopsis leaves plays a crucial role in carbon and nitrogen metabolism. *Plant Physiol.* 167, 671–681. doi: 10.1104/pp.114.254474

Skovbjerg, C. K., Angra, D., Robertson-Shersby-Harvie, T., Kreplak, J., Keeble-Gagnère, G., Kaur, S., et al. (2023). Genetic analysis of global faba bean diversity, agronomic traits and selection signatures. *Theor. Appl. Genet.* 136, 1–27. doi: 10.1007/s00122-023-04360-8

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., and Yoder, J. (2013). Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PloS One* 8, 65688. doi: 10.1371/journal.pone.0065688

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., et al. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science* 347. doi: 10.1126/science.1259855

Stoddard, F., and Hämäläinen, K. (2011). Towards the world's earliest maturing faba beans. *Grain Legumes.* 56, 9–10.

Torkamaneh, D., and Belzile, F. (2022). *Genome-Wide Association Studies*. Eds. D. Torkamaneh and F. Belzile (Quebec City, QC, Canada: Springer Nature). doi: 10.1007/978-1-0716-2237-7

Voisin, A. S., Guéguen, J., Huyghe, C., Jeuffroy, M. H., Magrini, M. B., Meynard, J. M., et al. (2014). Legumes for feed, food, biomaterials and bioenergy in Europe: A review. *Agron. Sustain. Dev.* 34, 361–380. doi: 10.1007/s13593-013-0189-y

Wang, J., Gao, S., Peng, X., Wu, K., and Yang, S. (2019). Roles of the INO80 and SWR1 chromatin remodeling complexes in plants. *Int. J. Mol. Sci.* 20, 4591. doi: 10.3390/ijms20184591

Wang, C., Liu, R., Liu, Y., Hou, W., Wang, X., Miao, Y., et al. (2021). Development and application of the Faba_bean_130K targeted next-generation sequencing SNP genotyping platform based on transcriptome sequencing. *Theor. Appl. Genet.* 134, 3195–3207. doi: 10.1007/s00122-021-03885-0

Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genom. Proteomics Bioinf.* 19, 629–640. doi: 10.1016/j.gpb.2021.08.005

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz, K., et al. (2016). A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba* L.). *Plant Biotechnol. J.* 14, 177–185. doi: 10.1111/pbi.12371

Wickham, H. (2016). ggplot2 Elegant Graphics for Data Analysis. 2. ed (Cham: Springer International Publishing).

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., et al. (2019). Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *Lancet* 393, 447–492. doi: 10.1016/S0140-6736(18)31788-4

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom. Proteomics Bioinf.* 19, 619–628. doi: 10.1016/J.GPB.2020.10.007

Zhang, C., Cao, L., Rong, L., An, Z., Zhou, W., Ma, J., et al. (2015). The chromatin-remodeling factor AtINO80 plays crucial roles in genome stability maintenance and in plant development. *Plant J.* 82, 655–668. doi: 10.1111/tpj.12840

Zhang, Y., Gu, L., Hou, Y., Wang, L., Deng, X., Hang, R., et al. (2015). Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates plant flowering by targeting alternative polyadenylation. *Cell Res.* 25, 864–876. doi: 10.1038/cr.2015.77

Zhao, Z., and Assmann, S. M. (2011). The glycolytic enzyme, phosphoglycerate mutase, has critical roles in stomatal movement, vegetative growth, and pollen production in Arabidopsis thaliana. *J. Exp. Bot.* 62, 5179–5189. doi: 10.1093/jxb/err223

Zhao, N., Xue, D., Miao, Y., Wang, Y., Zhou, E., Zhou, Y., et al. (2023). Construction of a high-density genetic map for faba bean (*Vicia faba* L.) and quantitative trait loci mapping of seed-related traits. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1201103

Zierer, W., Hajirezaei, M. R., Eggert, K., Sauer, N., von Wirén, N., and Pommerrenig, B. (2016). Phloem-specific methionine recycling fuels polyamine biosynthesis in a sulfur-dependent manner and promotes flower and seed development. *Plant Physiol.* 170, 790–806. doi: 10.1104/pp.15.00786

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2025:20

Faba bean holds promise for sustainable agriculture, yet breeding progress remains limited. This thesis integrates phenotypic, genetic, and biochemical analyses to explore agronomic traits and seed quality. A diversity panel of 220 accessions revealed significant trait variation, and a GWAS mapped 51 novel SNPs to key agronomic traits. Temporal analyses of seed development uncovered dynamic nutrient accumulation, while *PEBP* gene characterisation, including *VfTFL1a* variation, informs flowering and growth determinacy. These insights provide genomic resources for breeding faba bean in Northern regions.

Hannah Ohm received her graduate education at the department of Plant Breeding and her MSc degree in Molecular Biology with a specialisation in Plant and Forest Biotechnology from Umeå University, Sweden.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-455-0
ISBN (electronic version) 978-91-8046-505-2