

Institute of Plant Genetics

Polish Academy of Sciences

Marker-Assisted selection of white lupin (*Lupinus albus* L.)

Enhancing anthracnose resistance and reducing quinolizidine alkaloid content

Selekcja białego łubinu (*Lupinus albus* L.) wspomagana markerami

Zwiększenie odporności na antraknozę i zmniejszenie zawartości alkaloidów chinolizydynowych

Doctoral Dissertation

András Patyi

Supervised by

Assoc. Prof. Michał Książkiewicz, IGR PAN

Dr. Mariateresa Lazzaro. Research Institute of Organic Agriculture (FiBL)

May 2026

This doctoral dissertation was written following the regulations of the extramural mode procedure for doctoral proceedings at the Institute of Plant Genetics, Polish Academy of Sciences (Procedure II).

Major results of the research conducted within the doctorate were presented in the following two original publications:

1. Patyi, A.*, Kamp, M., Arncken, C., Biazzi, E., Książkiewicz, M., Messmer, M. M., Schneider, M., Tava, A., & Lazzaro, M. (2025). Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (Lupinus albus, L.) and development of ultra-low alkaloid recombinants by stacking with the pauper allele. BMC Plant Biology, 25(1), 945. https://doi.org/10.1186/s12870-025-06951-7
2. Patyi, A.*, Schneider, M., Arncken, C., Książkiewicz, M., Messmer, M. M., Schwertfirm, G., Lazzaro, M. (2026). From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments. *Mol Breeding* 46, 37. https://doi.org/10.1007/s11032-026-01661-w

*** corresponding author**

Both publications present original research published by peer-reviewed journals. The original findings from both publications, and their combined implication are summarized and presented in this dissertation. I am both the first and the corresponding author in these publications.

Funding

This dissertation and related published scientific articles received funding from the projects DIVINFOOD (funded by the EU, grant no.101000383), LiveSeeding (funded by the EU, SERI and UKRI, grant no.101059872) and LUPINNO SUISSE (funded by the Swiss Federal Office for Agriculture - FOAG, contract: 2020/51).

Furthermore, this doctoral study received funding from the project OPUS 23 project (grant no. 2022/45/B/NZ9/01397), funded by National Science Centre, Poland.



Acknowledgements

I would like to express my deepest gratitude to my supervisors, **Michał Książkiewicz** and **Mariateresa Lazzaro**, for their exceptional guidance throughout my PhD, their dedication to the subject, and for all that I have learned under their supervision.

This work is the amalgamation of the support and contributions of many people over the past four years. I am grateful to the members of the **Plant Breeding group** for providing an environment in which I could grow and flourish, to the **Gene Structure and Function team** for welcoming me, and to **FiBL** for their dedication and uplifting atmosphere.

My home away from home is shaped by the people around me, and I feel incredibly fortunate to have met so many fascinating individuals along the way. Your presence in my life means more to me than I can truly express. Thank you for your patience, your support, and for being part of this experience.

Végül, de nem utolsó sorban pár köszönő szót szeretnék címezni a szívemhez talán legközelebb álló szegletnek, a magyar buborékomnak. A szüleimnek, testvéreimnek és nagyszüleimnek azért, hogy egész életembe formálva egymást, jobb emberré tettek (remélem, hogy tettük egymást). Az ifjú-, felnőttkorom másik fontos formálói a barátaim. Elmondhatatlanul hálás vagyok azért, hogy ennyi idő után is egymás életének részei lehetünk. Általatok elhiszem, hogy jó ember vagyok. *Madarat tolláról, embert barátjáról*

"A hazám a házam tája, amely pillantásommal, ahogy növekszem, egyre bővül, széles gyűrűkben egyre távolabbi területekre terjed, mint a hullám a vízbe dobott kavics körül; világokat hódíthat, elérheti a csillagokat – amikor a régi ház már örökre elmerült."

Illyés Gyula, Puszták népe (1936)

"My native land is the region round my house, and as I grow up it gradually extends to include ever wider circles, like the ripples from a stone thrown into the water. It may overcome whole worlds and may even reach the stars, when the old home has disappeared for ever."

Illyés Gyula, People of the Puszta (1936, Translated by G. F. Cushing)

Table of Contents

Funding 2
Acknowledgements 3
Abbreviations 6
Summary 7
Summary in Polish - Streszczenie 9
Summary in Hungarian - Összefoglaló 11
Introduction 13
Aims 17
Research hypotheses 17
Alkaloid levels 17
Anthracnose resistance 18
Material and methods 19
Plant material 19
Panels for the QA experiment (Patyi et al., 2025) 19
Panels for the anthracnose experiment (Patyi et al., 2026) 19
Phenotyping 20
QA pheno- and chemotyping (Patyi et al., 2025) 20
Disease assessment (Patyi et al., 2026) 20
Sequencing (Patyi et al., 2025) 22
Genomic Prediction for anthracnose resistance (Patyi et al., 2026) 22
SNP marker assays: individual assessment and genomic prediction 22
Combined application across the two target traits 24
Results and discussion 25
Discovery of a novel low-QA determinant 25
Validation of markers tagging low-QA content and implications for breeding 25
Additive effects and biochemical consequences of allele stacking 26
Phenotypic variation in anthracnose resistance across panels ad environments 27
Limited transferability of marker associations tagging anthracnose resistance 27
Genomic prediction for anthracnose resistance: performance and constraints 27

Improving genomic prediction through retraining 28
Synthesis and application in breeding 29
Conclusion 31
Future prospects 31
References 33
Publication nr. 1 38
Supplementary information 54
Declaration of the PhD student's contribution in publications and the authors' declarations of their contribution 55
Publication nr. 2 65
Supplementary information 95
Declaration of the PhD student's contribution in publications and the authors' declarations of their contribution 96
Statement of thesis supervisors 105
Declaration of the author of the doctoral dissertation 105

Abbreviations

BBCH – Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie

BSA – Bulked Segregant Analysis

CAPS – Cleaved Amplified Polymorphic Sequence

dpi – days post inoculation

FiBL – Forschungsinstitut für Biologischen Landbau (Research Institute of Organic Agriculture)

GP – Genomic Prediction

GWAS – Genome Wide Association Study

gzpk – Getreidezüchtung Peter Kunz (Cereal Breeding Peter Kunz)

InDel – insertion-deletion

KASP – Kompetitive Allele Specific PCR

LfL - Bavarian State Research Center for Agriculture

LMM – Linear Mixed Model

MAS – Marker Assisted Selection

PACE – PCR Allelic Competitive Extension

PCR – Polymerase Chain Reaction

RIL – Recombinant Inbred Line

rrBLUP – Ridge Regression Best Linear Unbiased Predictor

sAUDPC – Standard Area Under Disease Progress Curve

SNP – Single Nucleotide Polymorphism

QA – Quinolizidine Alkaloid

Summary

White lupin (*Lupinus albus* L.) is a valuable grain legume, especially for sustainable agricultural systems both for its nutritious seeds, and for their beneficial ecosystem services during the growing season. However, its profitable production in temperate regions is hindered by two major challenges, susceptibility to anthracnose disease (caused by the plant-pathogenic ascomycete *Colletotrichum lupini*), as well as high and unstable quinolizidine alkaloid content in the seed. This dissertation explores the current *status quo* of marker assisted selection for these two traits of interest, and through diverse molecular analyses, present novel, functionally validated molecular markers.

Through Bulked Segregant Analysis on an F_2 population segregating for two low quinolizidine alkaloid (QA) loci, a novel quantitative trait locus (QTL) was identified on chromosome *Lalb_Chr05*. This locus represents a new genetic source of reduced alkaloid content, independent of the well-characterized and introduced into breeding *pauper* allele. By combining genotyping for the known *pauper* locus conferring the low-QA chemotype, together with the newly identified QTL, 50 recombinants carrying stacked alleles from both loci were identified in the subsequent generation (F_3). These lines showed exceptionally low total alkaloid content (22.8 ±10.4 ppm), substantially lower than even the *pauper*-associated genotypes (171.7 ±18.5 ppm).

Regarding anthracnose resistance in white lupin, nine previously published molecular markers were validated in multiple panels from various environments. Additionally, re-analysis of existing datasets enabled (i) the discovery of five additional single nucleotide polymorphism (SNP) markers through genome-wide association study (GWAS) and (ii) the development of a genomic prediction (GP) model based on 42 SNP loci. A previously reported QTL on chromosome *Lalb_Chr10*, associated with resistance in the cultivar 'Frieda' was positively validated. Two markers showed consistent statistically significant associations across multiple panels, while seven were panel-specific. The GP model achieved high prediction accuracy in the training population (0.84 ± 0.13), but required re-training for unrelated panels to maintain high predictive performance.

Together, these studies provide complementary molecular tools for further white lupin improvement as a crop plant. The identification of a novel low-alkaloid QTL and development of associated markers enable enhanced selection for seed quality, while validated SNP markers and genomic prediction approaches strengthen the breeding for anthracnose resistance. The study highlighted the value of combining established loci with newly discovered variation, as well as the need to validate molecular markers

in unrelated, and sometimes narrower genepools. These findings offer a robust integrated framework for accelerating the development of anthracnose-resistant white lupin varieties with high seed quality.

Summary in Polish - Streszczenie

Łubin biały (*Lupinus albus* L.) jest rośliną strączkową, wartościową szczególnie w kontekście rolnictwa zrównoważonego, zarówno ze względu wysoką wartość odżywczą nasion, jak i korzystny wpływ na ekosystem w okresie wegetacyjnym. Jednak jego opłacalna produkcja w regionach o klimacie umiarkowanym jest utrudniona przez dwa główne wyzwania: podatność na antraknozę (wywoływaną przez patogenicznego workowca *Colletotrichum lupini*) oraz wysoką i niestabilną zawartość alkaloidów chinolizydynowych w nasionach. Niniejsza rozprawa przedstawia aktualny stan selekcji wspomaganej markerami w odniesieniu do tych dwóch cech. Ponadto, wykorzystując różnorodne analizy molekularne, dostarcza nowe, zweryfikowane pod kątem funkcjonalności, markery molekularne.

Analiza segregantów zbiorczych (ang. Bulked Segregant Analysis) w populacji F_2 segregującej się pod względem dwóch loci o niskiej zawartości alkaloidów chinolizydynowych (QA) pozwoliła na zidentyfikowanie nowego locus cechy ilościowej (QTL) w chromosomie *Lalb_Chr05*. Stanowi on nowe źródło genetyczne obniżonej zawartości alkaloidów, niezależne od dobrze scharakteryzowanego i wdrożonego do hodowli allela *pauper*. Łącząc genotypowanie znanego locus *pauper*, nadającego chemotyp o niskiej zawartości QA, z nowo scharakteryzowanym QTL, zidentyfikowano, w następnym pokoleniu (F_3), 50 rekombinantów niosących skumulowane allele obu loci. Linie te wykazały wyjątkowo niską całkowitą zawartość alkaloidów (22,8 ±10,4 ppm), znacznie niższą nawet od genotypów typu *pauper* (171,7 ±18,5 ppm).

Biorąc pod uwagę kwestię odporności łubinu białego na antraknozę, dziewięć wcześniej opublikowanych markerów molekularnych zostało poddanych walidacji w kilku panelach pochodzących z różnych środowisk. Ponadto, ponowna analiza istniejących zbiorów danych umożliwiła (i) odkrycie pięciu dodatkowych markerów polimorfizmu pojedynczego nukleotydu (ang. single nucleotide polymorphism, SNP) w ramach badania asocjacyjnego całego genomu (ang. genome-wide association study, GWAS) oraz (ii) opracowanie modelu predykcji genomowej (ang. genomic prediction, GP) opartego na podstawie 42 loci SNP. Pozytywnie zweryfikowano opublikowany wcześniej QTL zlokalizowany w chromosomie Lalb_Chr10, związany z odpornością odmiany „Frieda". Dwa markery wykazały spójne, istotne statystycznie asocjacje w obrębie kilku paneli, podczas gdy siedem było specyficznych dla danego panelu. Model GP osiągnął wysoką dokładność prognozowania w populacji uczącej (0,84 ± 0,13), ale wymagał ponownego uczenia dla niezależnych paneli w celu utrzymania wysokiej wydajności prognostycznej.

Podsumowując, badania te dostarczyły uzupełniających się narzędzi molekularnych do dalszego udoskonalania łubinu białego jako rośliny uprawnej. Identyfikacja nowego QTL o niskiej zawartości alkaloidów i opracowanie zasocjowanych z nim markerów umożliwia ulepszoną selekcję pod kątem jakości nasion, podczas gdy potwierdzone markery SNP i narzędzia do prognozowania genomowego powinny ułatwić hodowlę pod kątem odporności na antraknozę. Przeprowadzone badania podkreśliły wartość połączenia analizy znanych loci z nowo odkrytymi wariantami, a także potrzebę walidacji markerów molekularnych w niezależnych, a czasem także węższych pulach genowych. Wyniki te oferują solidne, zintegrowane podstawy umożliwiające przyspieszenie opracowania odpornych na antraknozę odmian łubinu białego o wysokiej jakości nasion.

Summary in Hungarian - Összefoglaló

A fehér csillagfürt (*Lupinus albus* L.) egy ígéretes hüvelyes növény, amely magjainak tápanyagtartalma, és a vegetációs időszak alatt nyújtott jótékony ökoszisztéma-szolgáltatásai révén különösen alkalmas a fenntartható mezőgazdasági rendszerekben. Azonban, a mérsékelt éghajlati övezetekben történő jövedelmező termesztését két fő kihívás akadályozza: az antraknózisra (amelyet a *Colletotrichum lupini* nevű aszkomicéta okoz) való érzékenység, valamint a magok magas és ingadozó kinolizidin-alkaloid (KA) tartalma. Ez a disszertáció feltárja a két vizsgált tulajdonságra vonatkozó marker asszisztált szelekció jelenlegi helyzetét, és különböző molekuláris elemzések segítségével új, validált molekuláris markereket mutat be.

Először, egy alacsony kinolizidin-alkaloid tulajdonságra szegregáló F_2 populáción végzett Bulked Segregant Analysis módszerrel azonosítottunk egy új kvantitatív tulajdonságlókuszt (QTL) a *Lalb_Chr05* kromoszómán. Ez a lókusz az alacsony alkaloidtartalom új genetikai forrását jelenti, függetlenül a jól jellemzett *pauper* allélétől. Az alacsony KA-kemotípust biztosító ismert *pauper* lókusz, illetve az újonnan azonosított QTL-nek a kombinálásával genotipizáltuk a következő generációt, és azonosítottunk 50 olyan F_3 rekombinánst, amelyek egymásra épülő allélokat hordoznak. Ezek a vonalak kivételesen alacsony alkaloid-tartalmat mutattak (22.8 ± 10.4 ppm), ami lényegesen alacsonyabb volt még a *pauper*-hez köthető genotípusokénál is (171.7 ± 18.5 ppm).

Másodszor, a fehér csillagfürt antraknózis-rezisztenciájával kapcsolatban kilenc korábban publikált molekuláris markert validáltunk különböző környezeti körülmények mellett több panelben. Ezen felül a meglévő adathalmazok újbóli elemzése lehetővé tette (i) öt további SNP-marker felfedezését GWAS segítségével, valamint (ii) egy 42 SNP-n alapuló genomikus predikciós (GP) modell kidolgozását. Validáltuk a *Lalb_Chr10* kromoszómán található, a „Frieda" fajtában az ellenállással összefüggő, korábban jelentett QTL-t. Két marker több panelben is következetesen szignifikáns összefüggést mutatott, míg hét panel-specifikus volt. A GP-modell a tréning populációban magas predikciós pontosságot ért el (0.84 ± 0.13), de a prediktív teljesítmény fenntartása érdekében újraképzésre volt szükség a genetikailag távolabbi panelek esetében.

Ezek a tanulmányok egymást kiegészítő molekuláris eszközöket nyújtanak a fehér csillagfürt nemesítéséhez. Egy új, alacsony alkaloid-tartalmú QTL azonosítása és a kapcsolódó markerek kifejlesztése lehetővé teszi a magminőség javítását célzó szelekciót, míg a validált SNP-markerek és a genomikus predikciós megközelítések erősítik az antraknózis-rezisztencia nemesítését. Mindkettő rávilágít a már ismert

lókuszok és az újonnan felfedezett variációk kombinálásának értékére, valamint arra, hogy a molekuláris markereket független, és néha szűkebb génállományokban is validálni szükséges. Ezek az eredmények szilárd, integrált keretrendszert nyújtanak a kiváló minőségű, antraknózis rezisztens fehér csillagfürt fajták nemesítésének előresegítéséhez.

Introduction

White lupin (*Lupinus albus* L.) is a Mediterranean grain legume used both for its highly nutritious seeds and its beneficial contributions to the environment during the growing season. White lupin biomass has been used for multiple purposes, such as green manure and fodder, and the harvested seeds can be utilized as feed and food. During the growing season, white lupin represents an important niche, since like most legumes, they host proteobacteria (collectively known as rhizobia) in their root nodules, able to fix atmospheric nitrogen. Two α-proteobacterial genera were described to live in symbiotic associations in white lupin root nodules, namely *Bradyrhizobium* and *Ochrobactrum*, the latter being less effective. Due to its symbiosis with *Bradyrhizobium* spp., white lupin can fix ca. 150-200 kg of dinitrogen per ha, which dramatically reduces the amount of fertilizer that should be used (Barrera et al., 1997; Robinson et al., 2000). Concerning another important macroelement, phosphorous, white lupin represents a niche even amongst fellow legumes. By forming cluster roots in response to phosphate deficiency, white lupin is able to mobilize soil phosphates which process further improves the content of organic matter in soils (Xu et al., 2020). This means that potassium is the only macroelement fertilizer potentially required during white lupin production (Van De Noort, 2017), making this crop especially advantageous for low-input farming systems.

Apart from green manure and fodder, white lupin is produced for its seed, having high protein (33-47% depending on genotype and location) and excellent fiber content (Huyghe, 1997). Furthermore, white lupin seeds have an excellent fatty acid composition ratio of omega-3 to omega-6, they contain non-starch carbohydrates, oligosaccharides and antioxidants making them a great addition to a healthy diet (Prusinski, 2017). However, an important aspect for consumption of white lupin is the alkaloid content of the seeds. Seeds (dry weight) of wild forms of white lupin could contain up to 2.2% of toxic quinolizidine alkaloids (Kroc et al., 2017). Historically, the grains were soaked in running water to leech out the quinolizidine alkaloids, but nowadays bitterness-lacking (sweet) varieties exist with the alkaloid content ranging from 0.01-0.08% of seed dry weight (Words, 1982). This ratio of the sweet varieties are below the suggested threshold of alkaloid content for human consumption of 0.02% (Bundesinstitut Für Risikobewertung, 2017; Schrenk, Bodin, Chipman, del Mazo, Grasl-Kraupp, et al., 2019), a major achievement of previous breeding efforts.

The formal breeding of white lupin started in the early 20th century with the work of Prof. Reinhold von Sengbusch, who worked extensively with lupins. His work was a prime example of exploiting Mendelian inheritance for directional selection against common wild type traits such as pod shattering and -shedding, non-permeable seed

coat, late flowering, dark seeds and high alkaloid content (Hondelmann, 1984). In most varieties with low alkaloid content, the allele *pauper* confers the trait of sweetness, although there are less exploited alleles, known such as *nutricius*, *exiguus*, *mitis*, and *reductus* representing possible recessive mutations (Words, 1982). However, these lesser alleles have been reported during the mid-20th century based on crossings, thus not much is known about the transcriptomic regulation and genes involved conferring low alkaloid content in white lupins. Two studies presented markers tagging this trait, with the reported ability to distinguish between *pauper*, *exiguus* and *nutricius* (Lin et al., 2009; Rychel & Książkiewicz, 2019). Later, with the reference- and pangenome of white lupin, Hufnagel et al. (2021) suggested another quantitative trait locus (QTL), indicating that these loci could be associated to other, non-*pauper* mutations previously mentioned. Years later, the causal mutation conferring the sweetness associated to *pauper* was described by Mancinotti et al. (2023). They identified the gene on chromosome *Lalb_18* (*Lalb_ Chr18g0051511*), and associated SNP, that carries the causal mutation that is responsible to an early change in the biosynthetic process of quinolizidine alkaloids (QAs), resulting in low alkaloid content. However, despite this major achievement, seeds harvested from reportedly low QA *pauper* accessions are exceeding the suggested threshold (200 mg/kg dry weight(Schrenk, et al., 2019)) for consumption (Kroc et al., 2017; Schryvers et al., 2023; FiBL, unpublished data), maintaining the need for stably sweet varieties, especially across growing seasons. Furthermore, unveiling genomic regions associated to a sweetness source other than *pauper* would lead to the possibility to stacking these low QA conferring loci, reaching closer towards of ultra-sweet white lupins.

The other challenge in white lupin cultivation was initiated by one of the crop`s most devastating pathogens, *Colletotrichum lupini* (Figure 1A), causing up to total yield losses in unfavorable years (Alkemade et al., 2025). This hemibiotrophic ascomycete, causing anthracnose disease in all lupins appeared in the 1970s and 80s in Europe and South America, quickly becoming a worldwide pandemic (Gondran et al., 1999; Talhinhas et al., 2016). Concerning plant protection, it is an extremely difficult disease to manage due to multiple factors, the first one being the source of infection. Primary infection originates from infected seeds. Reportedly, as low as 0,01-0,1% of infected seeds sown lead to severe yield losses (Sreenivasaprasad & Talhinhas, 2005; Thomas & Sweetingham, 2004). Secondary infection occurs due to the spreading of conidia, infecting through rain splash (Alkemade et al., 2025). In either case, the disease progression is swift, after a short biotrophic phase (up to four days) causing stem- and petiole twisting (Figure 1B), the fungus enters into necrotrophy (Figure 1C) killing its host, with reported formation of conidia 6 days onwards after infection (Alkemade et al., 2025; Guilengue et al., 2022). In white lupin, resistance to anthracnose has been

described to be polygenically controlled (Alkemade et al., 2025; Hufnagel et al., 2020; Książkiewicz et al., 2017), making resistance breeding challenging.

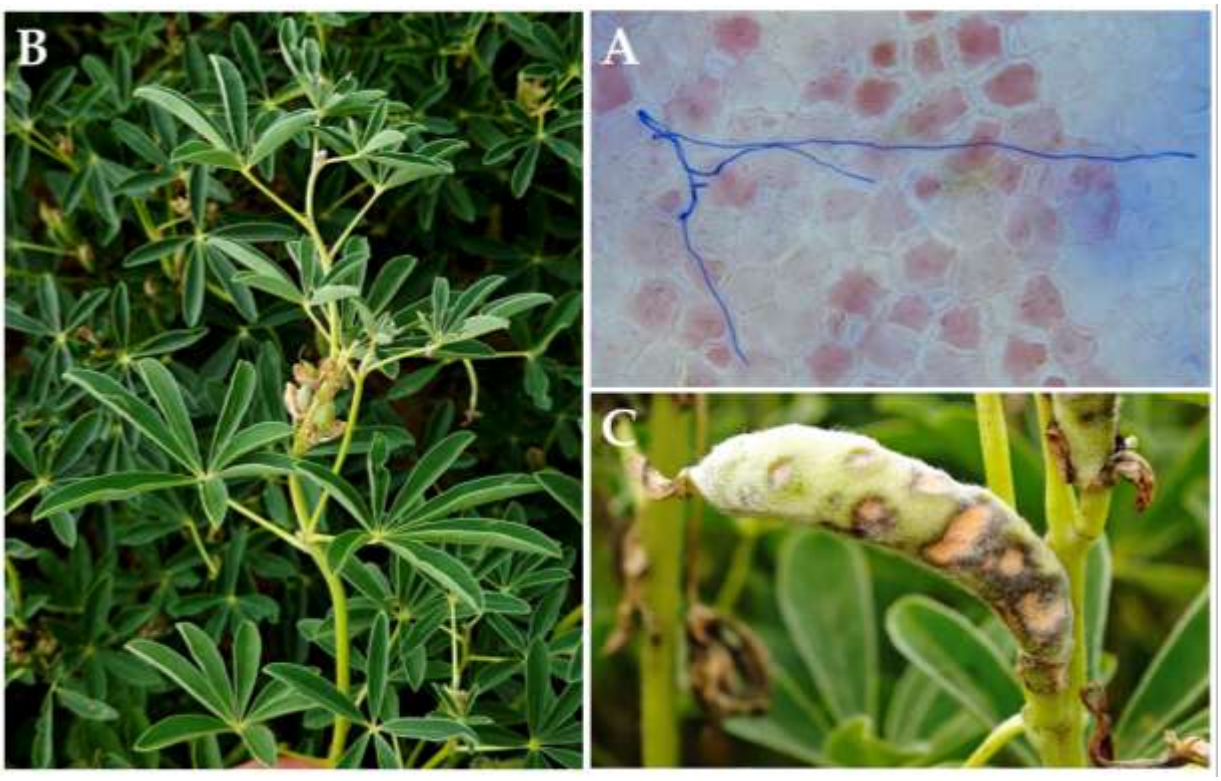


Figure 1. Typical symptoms of *Colletotrichum lupini* on white lupin. (**A**) Fungal hyphae growing out of a stoma 3 days post inoculation (fungal tissue stained with lactophenol cotton blue, cleared plant tissue with safranin-O). (**B**) Typical symptom in early- to mid-season including stem twisting and wilting (field, 2024). (**C**) Typical symptom in late-season including sporulating lesions on all aerial parts of the plant (field, 2024)

Multiple studies explored through different molecular approaches determinants of anthracnose resistance in white lupin and suggested functionally validated molecular markers associated to them. However, there is limited transferability of information between the studies throughout the years, lacking validation in unrelated material, and no reported application within other breeding programmes.

This dissertation aims to lay the molecular groundwork for stably sweet, anthracnose resistant white lupin. Through Bulked Segregant Analysis, we identified a novel locus associated to low alkaloid content and developed a molecular marker tagging this source. Combining the genotyping for *pauper*, we selected 50 F_3 stacked allele recombinants, showing ultra-low QA content with reduced QA profile. Subsequently, regarding anthracnose resistance, we validated the transferability of published and newly described molecular markers both in unrelated materials, with phenotypic values of the validation panels from controlled conditions and from field observations. We proposed a functionally validated a minimal set of 42 SNP markers,

intended for genomic prediction (GP), which showed high accuracy in the training panel, but initially lacked in performance in unrelated material. However, re-training the GP model with the validation data significantly improved predictive power, revealing potential for transferability in narrower and unrelated genepools.

One of the panels used for validation consisted of the aforementioned stacked allele recombinants, in which re-training the GP model resulted in an improved prediction accuracy of 0.62 ± 0.11 for anthracnose resistance. The results from the prediction model, together with the phenotypic data originating from the field highlights the value of this segregating population. These SNP markers were proven valuable for marker assisted selection in white lupin, and the segregating population represent the success of their application in an active breeding programme.

Aims

The general aim of this PhD study was the validation of previously described loci and the identification of additional ones associated to major traits of interest in white lupin to improve marker-based selection application in breeding this neglected crop.

The specific objectives of the study were:

- Identifying regions yet to be described conferring low-QA chemotype through Bulked Segregant Analysis;
- Developing and evaluating molecular markers tagging a novel source of the low-QA chemotype;
- Validating existing and newly described individual molecular markers tagging anthracnose resistance;
- Develop markers for genomic prediction for anthracnose resistance, validated in diversity- and breeding panels.

Research hypotheses

Alkaloid levels

- The sweetness of the 'Dieta' may be conferred by a non-*pauper* locus (one of the previously reported less exploited alleles or a new one) because this accession has low QA level but shows a bitter *pauper* genotype when analyzed by diagnostic *pauper* markers
- If the genetic background of Dieta sweetness is different than *pauper*, it should be possible to develop segregating population by crossing it with a *pauper* donor as well as to obtain stacked allele recombinants accumulating potentially lower QA levels than those carrying single loci (however the latter may depend on the molecular function of the non-*pauper* locus).
- Such a segregating population can be later exploited for identification of SNPs associated with a non-*pauper* locus, potentially useful for further tracking

desired allele(s) in the offspring where these low-QA carrying alleles are present

Anthracnose resistance

- The trait of anthracnose resistance is polygenically controlled in white lupin. Several studies reported different loci associated with anthracnose resistance, usually located in different genomic regions, what indicates low reproducibility of QTLs between different studies, resulting potentially from high complexity of quantitative control of the resistance, strong environmental effects and/or significant genotype x environment interactions. Most often, molecular markers are functionally validated in interset- and progeny selection approach, limiting transferability.
- Rapid linkage disequilibrium decay reported for white lupin may make finding universal selective markers difficult or even impossible. In such a case GP may be more suitable approach, as it is anchored in numerous loci distributed across the whole genome, whereas QTL markers target just a few loci. To make GP approach feasible with PCR-based technology, it should constitute a relatively short list of PCR-applicable SNPs providing reasonable predictive ability. We assume that such a list may be achieved from the large SNP dataset by iterative marker thinning with appropriate statistical support.

Validation of markers and the GP model across related and unrelated panels will provide information about their potential applicability for genomic selection. We anticipate that improvement may be achieved by retraining of the GP model, when working with unrelated material to the training population.

Material and methods

The materials and methods used for the two studies are presented in the original publications (Patyi et al., 2025, 2026). The following section is a shortened summary of the materials and methods used in these studies.

Plant material

Panels for the QA experiment (Patyi et al., 2025)

Two commercial white lupin cultivars, 'Frieda' (*pauper* sweetness) and 'Dieta' (unknown, non-*pauper* sweetness), were crossed in 2021. F_1 was selfed to produce F_2 seeds. The 112 individuals of the F_2 population (Table 1, Patyi et al., 2025, Supplementary Table S1) were grown in 2023 in field conditions under protective nets to avoid cross-pollination, and leaf tissue was collected for downstream analyses. From this population, 13 heterozygous individuals at the *pauper* locus were selected to establish an F_3 population (Table 1), comprising 329 plants grown in 2024 under similar field conditions (Patyi et al., 2025, Supplementary Table S3). Parental lines were also included as controls.

Additionally, a diverse validation panel (Table 1) of 34 genotypes—including breeding lines, cultivars, landraces, and a wild type—was assembled alongside representatives of the 39 accessions from the white lupin pangenome (Hufnagel et al., 2021) to validate the pauper-tagging and the novel sweetness source-tagging molecular markers (Patyi et al., 2025, Supplementary Table S5).

Panels for the anthracnose experiment (Patyi et al., 2026)

Publicly available datasets were used (Table 1), including the LUW panel (Schwertfirm et al., 2024) consisting of 255 accessions (gene bank material, breeding lines, and cultivars) and a second panel of 200 accessions (192 with sufficient phenotypic data) consisting of cultivars, landraces, and breeding lines (Alkemade et al., 2022).

Additional genetic resources were curated to evaluate marker performance and genomic prediction accuracy. These included (i) a diversity panel of 48 accessions representing broad genetic backgrounds (Table 1, Patyi et al., 2026, Supplementary Table S1), (ii) a multi-parental panel of 96 individuals derived from advanced breeding lines and control cultivars (Table 1, Patyi et al., 2026, Supplementary Table S2, Supplementary Table S3), and (iii) a biparental F_3 population (50 individuals) from a 'Frieda' × 'Dieta' cross (Table 1, Patyi et al., 2026, Supplementary Table S4) consisting of stacked allele recombinants for the low-QA chemotype.

Phenotyping

QA pheno- and chemotyping (Patyi et al., 2025)

Alkaloid content in the F_2 and F_3 populations was initially assessed in the field using the Dragendorff paper test (Patyi et al., 2025, Figure 1.), which provides a semi-quantitative estimate based on colour intensity of plant sap reactions, dependent on QA content. In the F_2 population, bitterness was additionally evaluated by tasting flowers and scored on a 1–9 scale (sweet-bitter), which score was combined with the Dragendorff score (1-4, sweet-bitter) to generate an overall alkaloid-related phenotype. These samples, together with the parental plants were used for the Bulked Segregant Analysis.

For precise quantification, selected F_3 seeds and parental lines were chemically analysed. Quinolizidine alkaloids were extracted from milled seed material, purified, and quantified using gas chromatography with flame ionization detection (GC-FID), with sparteine as an internal standard. Compound identification of specific QAs was further confirmed by gas chromatography–mass spectrometry (GC-MS).

All statistical analyses (except sequencing data) were performed in R version 4.4.2 (R Core Team, 2024). Phenotypic data from the F_2 and F_3 populations were analysed using linear mixed models (package *lme4* version 1.4–36, Bates et al., 2015), with allelic combinations at candidate loci included as fixed effects and field-related factors as random effects.

Both semi-quantitative (Dragendorff scores) and quantitative traits (total alkaloid content and alkaloid ratios) were evaluated, and estimated marginal means were used to test for significant differences ($p < 0.05$) and assess genotype–phenotype associations, using the package *emmeans* version 1.10.7 (Lenth, 2017).

Disease assessment (Patyi et al., 2026)

Phenotyping for anthracnose resistance was conducted under both field- and controlled conditions across different panels and years. Field trials (2024) included the multi-parental (Patyi et al., 2026, Supplementary Table S2) and biparental panels (Patyi et al., 2026, Supplementary Table S4) grown under high infection pressure, with disease severity recorded of individual plants. In the multi-parental panel, repeated scoring enabled calculation of the standard area under disease progression curve (sAUDPC) value, while the biparental panel was assessed once late-season to reflect practical early-generation selection.

Controlled environment experiments involved artificial inoculation of 6 individuals from the diversity panel (Patyi et al., 2026, Supplementary Table S1) and the progeny from the field of the multi-parental panel (Patyi et al., 2026, Supplementary Table S3). 2 week-old plants were inoculated using stem-wound inoculation with the virulent *Colletotrichum lupini* JA01 strain (Alkemade, Messmer, Arncken, et al., 2021; Alkemade, Messmer, Voegele, et al., 2021), followed by repeated disease assessments on a 1-9 scale (healthy-dead, as described by Alkemade, Messmer, Arncken, et al., 2021) to calculate sAUDPC values. This phenotypic data was used to validate the correlation between disease scores made on the field and under controlled conditions, as well as to assess marker performance.

Phenotypic data from controlled experiments were analysed using linear mixed models (package *lme4* version 1.4–36, Bates et al., 2015) to estimate adjusted means (using *emmeans* version 1.11.1, Lenth et al., 2025), while field data were analysed at the individual plant level. Broad-sense heritability (H^2) was estimated for controlled datasets, whereas field heritability could not be determined due to lack of replication.

Table 1. Overview of all white lupin panels used in this dissertation. Marking if the dataset was publicly available or if it was generated for one of the two studies presented, the trait of interest, the purpose, the phenotyping environment and the type of phenotypic data generated from each panel.

Panel	Status	n	Trait	Purpose	Environment	Phenotyping
LUW	public	255	anthracnose	GWAS	field	Average disease score
Training	public	192	anthracnose	GP	climate chamber	sAUDPC
Multi-parental	*de novo*	22	anthracnose	Validation	climate chamber	sAUDPC
Multi-parental	*de novo*	96	anthracnose	Validation	field	sAUDPC
Diversity	*de novo*	48	anthracnose	Validation	climate chamber	sAUDPC
Biparental F_2	*de novo*	112	alkaloid	Bulked Segregant Analysis	field	Dragendorff
Biparental F_3	*de novo*	329/50	alkaloid / anthracnose	Validation	field	Dragendorff + GC-MS / disease score

Validation	*de novo*	34	alkaloid	Validation	climate chamber	Dragendorff

Sequencing (Patyi et al., 2025)

Bulked segregant analysis was performed using DNA from the F_2 population derived from the 'Dieta' × 'Frieda' cross. Two bulks were created from eight individuals each, representing sweet and bitter extremes of the phenotypes, based on a combined alkaloid-related score. Individuals homozygous for the *pauper* locus were excluded to focus on alternative genetic determinants of low alkaloid content. In addition, parental bulks (eight plants per cultivar) were included for sequencing, providing reference profiles for comparison.

For the Bulked Segregant Analysis, whole genome sequencing of parental and phenotypic bulks was performed at Novogene UK (Cambridge, UK) with a NovaSeq X Plus Series, paired-end technology (15× coverage, 150 bp long reads). Raw reads were quality-checked, aligned to the reference genome, and filtered to remove duplicates, low-quality reads, and ambiguous mappings as described by Schneider et al. (2022). Variant calling focused on previously reported SNPs to increase reliability, followed by stringent quality filtering.

Gene-associated SNPs were used to construct haplotypes, and allele frequencies were calculated for sweet and bitter bulks. These data were analysed using a Bayesian approach using *Baypass* (Gautier, 2015) to detect genomic regions under selection, and results were visualized using Manhattan plots.

Genomic Prediction for anthracnose resistance (Patyi et al., 2026)

Genomic prediction was performed in R version 4.2.1 (R Core Team, 2022) using the rrBLUP framework (version 4.6.1, Endelman, 2023) on a published dataset (Alkemade et al., 2022), with missing genotypic data imputed prior to analysis. Model performance was evaluated through extensive cross-validation across varying training set sizes and phenotypic inputs, with sAUDPC selected as the primary trait.

To optimize marker efficiency, iterative marker thinning was conducted based on effect size, prediction accuracy, and statistical significance, resulting in a reduced set of 50 informative SNPs. Due to technical limitations in assay design, a final set of 42 SNP markers was retained and used to build the genomic prediction model. The selected markers were used to assess population structure via principal component analysis (PCA) using *PCAtools* version 2.14.0 (Blighe & Lun, 2023).

SNP marker assays: individual assessment and genomic prediction

For all SNP marker assay, DNA was extracted using FTA cards (QIAGEN GmbH; Hilden, Germany), and polymorphic loci were genotyped using PACE assays (3CR Bioscience; Essex, UK). Allele-specific primers were designed by 3CR Bioscience (Essex, UK), with either alleles associated to a different fluorophore. PCR amplification was performed under established conditions using the CFX Opus 384 Real-Time PCR System (Bio-Rad Laboratories, Inc.; Hercules CA, USA), and allele calls were assigned automatically using the CFX Maestro version 2.2 (Bio-Rad Laboratories, Inc.; Hercules CA, USA) software. The resulting genotypic data were used for downstream analyses, including marker evaluation and genomic prediction.

PACE assays with allele-specific fluorescent primers associated with alkaloid content, including the *pauper* locus and markers within a newly identified QTL are presented in Patyi et al. (2025), Table 1. The F_2 population was initially genotyped for the *pauper* marker and later for the additional locus, while the F_3 population was genotyped for both markers. Similarly, PACE assays were developed for all markers associated with anthracnose resistance. Individually assessed loci originate from previously published and a *de novo* GWAS (Patyi et al., 2026, Table 2), in addition to 42 loci that represented the marker set used for genomic prediction (Patyi et al., 2026, Table 3).

All individually assessed marker were analysed in R 4.4.2 (R Core Team, 2024). Both semi-quantitative (Dragendorff scores), quantitative alkaloid scores were evaluated using LMMs (package *lme4* version 1.4–36, Bates et al., 2015), with genotype included as a fixed effect and experimental structure as a random effect. Estimated marginal means (calculated using *emmeans* version 1.11.1, Lenth et al., 2025) were used to test for significant differences and to quantify genotype–phenotype associations. Regarding individually assessed markers tagging anthracnose resistance, allelic states at QTL- and GWAS-derived loci associated to anthracnose resistance were analysed ANOVA or t-tests, depending on the number of allelic groups. The contribution of each marker to phenotypic variation was quantified by calculating the proportion of individual explained variance. To evaluate the combined performance of associated markers, phenotypic variance explained (PVE) was calculated across validation panels using an adjusted R^2 approach, considering only polymorphic markers in each dataset.

The performance of GP model for anthracnose resistance was assessed in both training and independent validation panels using cross-validation, evaluating prediction accuracy through correlation, root mean square deviation (RMSE), and coefficient of determination (R^2) as metrics. Additional models were generated by retraining by combining the training and respective validation panels to improve predictive performance.

Combined application across the two target traits

The 50 stacked allele recombinants were used in both studies, thus allowed merging the findings for application. A composite selection index was calculated to integrate multiple phenotypic and predicted traits into a single breeding value for visualization and comparison. The index combined (i) Dragendorff score (QA; scale in the population 1–3), (ii) observed disease score (R_obs; scale in the population 4–9), and (iii) predicted disease score (R_pred; scale in the population 5.54–6.89).

To ensure comparability across traits with different scales, all variables were linearly normalized to a 0–1 range prior to integration. The combined index was then calculated as a weighted sum of the normalized traits. The weighting scheme was chosen *a priori*:

$$Combined\ Index = 0.2 * QA + 0.5 * R_obs + 0.3 * R_pred$$

Observed disease resistance was given the highest weight (0.5), as it represents the directly measured phenotype and is therefore considered the most biologically relevant indicator of resistance. The predicted phenotypic score (0.3) was included to capture the underlying genetic potential, while Dragendorff score (0.2) contributed additional trait information of secondary importance, as all individuals in the population were phenotypically sweet.

Results and discussion

The results of the two independent studies are presented in the original publications (Patyi et al., 2025, 2026). This section contains a summary of the most important results from each publication, their synthesis, and one combined application aspect.

Discovery of a novel low-QA determinant

The phenotypic segregation ratio in the F_2 population is expected to follow a 1:3 (sweet:bitter) pattern if the *pauper* locus is the only determinant. Instead, we observed a ratio of 7:9, suggesting another determinant. Together with the observed 1:2:1 genotypic segregation of alleles at *pauper*-associated SNP locus (Lalb_Chr18_12359687), our results supports the presence of at least one additional, independently segregating recessive determinant associated to low quinolizidine alkaloid content (Patyi et al., 2025, Table 2). The absence of segregation distortion further supports the Mendelian inheritance of at least two loci controlling the trait.

Bulked Segregant Analysis (BSA) identified a major QTL on chromosome *Lalb_Chr05* associated with low alkaloid content. Despite relatively small bulk sizes (n = 8 per bulk), a clear signal was detected (Patyi et al., 2025, Fig. 2), highlighting the robustness of BSA even under constrained sampling. The identified QTL spans a relatively large genomic region, likely reflecting the limited number of individuals in the phenotypic extremes. Within this region, ten candidate genes were selected, including a locus encoding a putative aminoacyltransferase (*Lalb_Chr05g0222381*), a similar function to the gene associated to *pauper* (*Lalb_Chr18g0051511*). While this gene represents a biologically plausible candidate (thus was chosen for marker selection), its causal role could not be confirmed in the present study.

Validation of markers tagging low-QA content and implications for breeding

The candidate region (*Lalb_Chr05g0222381* ± 1000 bp) contained 54 polymorphic SNPs when comparing 'Dieta' (low alkaloid content from Lalb_Chr05 locus) and 'Frieda' (low alkaloid content from *pauper* locus on Lalb_Chr18). However, when expanding the data on the allelic states in all available sequenced accessions, only three loci were polymorphic and associated to the assumed alkaloid status within the pangenome (Hufnagel et al., 2021). These three SNPs were evaluated in the validation panel (Patyi et al., 2025, Supplementary Table S5), and the SNP Lalb_Chr05_6643621 was selected as candidate marker for validation in the F_2 and F_3 breeding populations, alongside the causal SNP associated to the *pauper* locus (Lalb_Chr18_12359687). Although, inconsistent marker-trait association of a few accessions from the validation panel

(Patyi et al., 2025, Supplementary Table S5) suggests that it is likely linked rather than causal.

In the F_3 population, haplotypes at the newly validated locus on chromosome *Lalb_Chr05* and the one linked to *pauper* were complemented with the Dragendorff phenotypic score, identifying 50 stacked allele recombinants. Representatives of haplotype groups were selected and sent for analytical quantification of QAs to measure total QA content and QA profile associated to each genotypic groups (Patyi et al., 2025, Table 3).

Additive effects and biochemical consequences of allele stacking

Both loci independently reduced alkaloid content compared to the wild type phenotype, but their combination resulted in a strong additive effect (Patyi et al., 2025, Table 3). Stacked allele recombinants exhibited extremely low total QA levels (as low as ~23 ppm on average), far below previously reported values in white lupin germplasm and well under safety thresholds for consumption of 200 ppm (Patyi et al., 2025, Fig. 3). This represents a substantial advancement in reducing alkaloid content in white lupin breeding.

At the biochemical level, distinct alkaloid profiles were associated with each genetic determinant (Patyi et al., 2025, Fig. 5). The *pauper* locus, known to block the biosynthetic pathway early, resulted in the absence of compounds such as ammodendrine (Mancinotti et al., 2023). In contrast, the chromosome *Lalb_Chr05* determinant altered downstream metabolic steps, leading to accumulation of specific intermediates and derivatives. When combined, these loci produced a highly simplified alkaloid profile, retaining only lupanine and a few esterified derivatives.

These observations suggest that the chromosome *Lalb_Chr05* determinant acts within the branch of the pathway converting lupanine to its hydroxylated derivatives. Changes in the ratio between 13α-hydroxylupanine and lupanine further support this hypothesis, indicating a shift in metabolic flux (Patyi et al., 2025, Fig. 6). However, the exact causal gene remains unresolved. Undoubtedly, this SNP locus is linked to the low-QA chemotype, however, the rapid linkage disequilibrium decay in white lupin (Hufnagel et al., 2021) leaves the possibility that the published marker may not be tightly linked to the causal variant. Nonetheless, based on the marker-trait association, the simultaneous selection in the present context for the novel locus on *Lalb_Chr05* and for *pauper* was validated by the successful identification of the stacked allele recombinants.

While the extremely low alkaloid content achieved is desirable for consumption, it may have ecological and agronomic implications. For example, reduced alkaloid levels

have been associated with increased susceptibility to pests such as aphids (Philippi et al., 2015, 2016), highlighting the need to further evaluate the performance and resilience of these genotypes under field conditions, where this issue is highly relevant.

Phenotypic variation in anthracnose resistance across panels ad environments

Across all evaluated panels, no accession exhibited complete resistance to anthracnose, confirming the quantitative nature of the trait. However, substantial variation was observed, with continuous distributions of resistance levels in both diversity and breeding panels. Some breeding lines outperformed the commercial cultivar 'Frieda', indicating the potential of the ongoing genetic improvement. Heritability estimates under controlled conditions were moderate to high, confirming a strong genetic component in that environmental context (Patyi et al., 2026, Fig. 1). However, correlations between field and controlled conditions were non-significant ($p = 0.718$) in the multi-parental panel where data from both environments was available, suggesting strong environmental effects and genotype-by-environment interactions. This may also be caused by the phenological stage, since 2 weeks-old seedling were used for phenotyping under controlled conditions, while adult plants were screened on the field. This also addresses potential concerns for transferability of results between different studies, and urges validation of previously published markers.

Limited transferability of marker associations tagging anthracnose resistance

Markers previously reported in the literature showed reduced effect sizes and inconsistent performance outside their original populations (Patyi et al., 2026, Fig. 2). This limited transferability can be attributed to differences in genetic background, overrepresentation of specific resistance sources (e.g., 'Frieda'-derived resistance), and environmental influences and differences in phenological stage. However, one marker (Lalb_Chr10_16065293) located in a previously identified QTL region was significant in the biparental panel, supporting its relevance tagging resistance associated to the parental cultivar `Frieda`, but also highlighting context dependency.

From the 14 SNP markers that were evaluated (5 from the re-analysis of an existing dataset, and 9 from literature), 8 SNPs had significant associations in at least one panel (Patyi et al., 2026, Fig. 3). The most consistent marker (Lalb_Chr06_2929006) from the repeated GWAS had significant associations in three of the four validation panels. Regardless, no single marker was universally successful, reinforcing the complexity and population specificity of resistance loci.

Genomic prediction for anthracnose resistance: performance and constraints

A genomic prediction model based on a minimal set of 42 SNP markers achieved high prediction accuracy ($r_{original}$ = 0.84 ± 0.13) within the training dataset, demonstrating even such a reduced marker set can capture substantial genetic signal under controlled conditions. However, predictive performance declined sharply when applied to independent panels, with low and unstable correlations and frequently negative coefficients of determination (Patyi et al., 2026, Fig. 5).

This loss of accuracy reflects several key limitations: reduced allelic diversity in breeding panels (Patyi et al., 2026, Fig. 4), differences in phenotyping environments, and genotype-by-environment interactions. Even in the diversity panel, which was genetically closest to the training population, prediction accuracy remained low and highly variable ($r_{original}$ = 0.30 ± 0.33). These findings are consistent with other studies showing reduced performance of genomic prediction across populations and environments, particularly for complex traits.

Marker density also plays a critical role. While reducing marker numbers can minimize noise, excessive thinning may eliminate informative loci, especially in species with rapid linkage disequilibrium decay such as white lupin. The balance between marker quantity and relevance remains a key factor.

Improving genomic prediction through retraining

Retraining genomic prediction models with combining the training dataset with a respective validation panel greatly improved previously unsuccessful prediction accuracy, particularly when field-derived phenotypes were included. Prediction accuracy increased substantially in the multi-parental ($r_{improved}$ = 0.66 ± 0.09) and biparental ($r_{improved}$ = 0.62 ± 0.11) panels after retraining, demonstrating that incorporating relevant genetic diversity and additional phenotypic values enhances model applicability. However, improvements remained context-dependent and limited by sample size and panel composition. Prediction accuracy of models retrained with phenotypic data from controlled-conditions improved compared to the original, but remained unstable.

These results highlight that genomic prediction models are most effective when applied to subset-, interset- or progeny selection and when phenotyping conditions are comparable. However, this introduces the limitation to apply available markers from literature to an unrelated population. The presented re-training strategy overcomes the challenge to integrate markers derived from controlled and field-based studies, capturing environment-specific effects and improving robustness. This represents a bridge between independent studies and breeding programmes, allowing transferability and application.

Synthesis and application in breeding

Together, our results highlight contrasting genetic architectures of two key traits in white lupin. Alkaloid content is controlled by loci with observably strong additive effects, enabling efficient marker-assisted selection and rapid genetic gain. In contrast, anthracnose resistance is polygenic, limiting the utility of individual markers and favoring genomic prediction approaches.

The marker located on Lalb_Chr05 enabled the identification of individuals in the biparental F_3 population carrying favorable alleles in both low-QA associated loci. In parallel, the available disease score values allowed the determination of the combined index of these individuals. Figure 2 illustrates the relationship between observed and predicted disease scores, with individuals color-coded by a composite index, integrating the alkaloid phenotype.

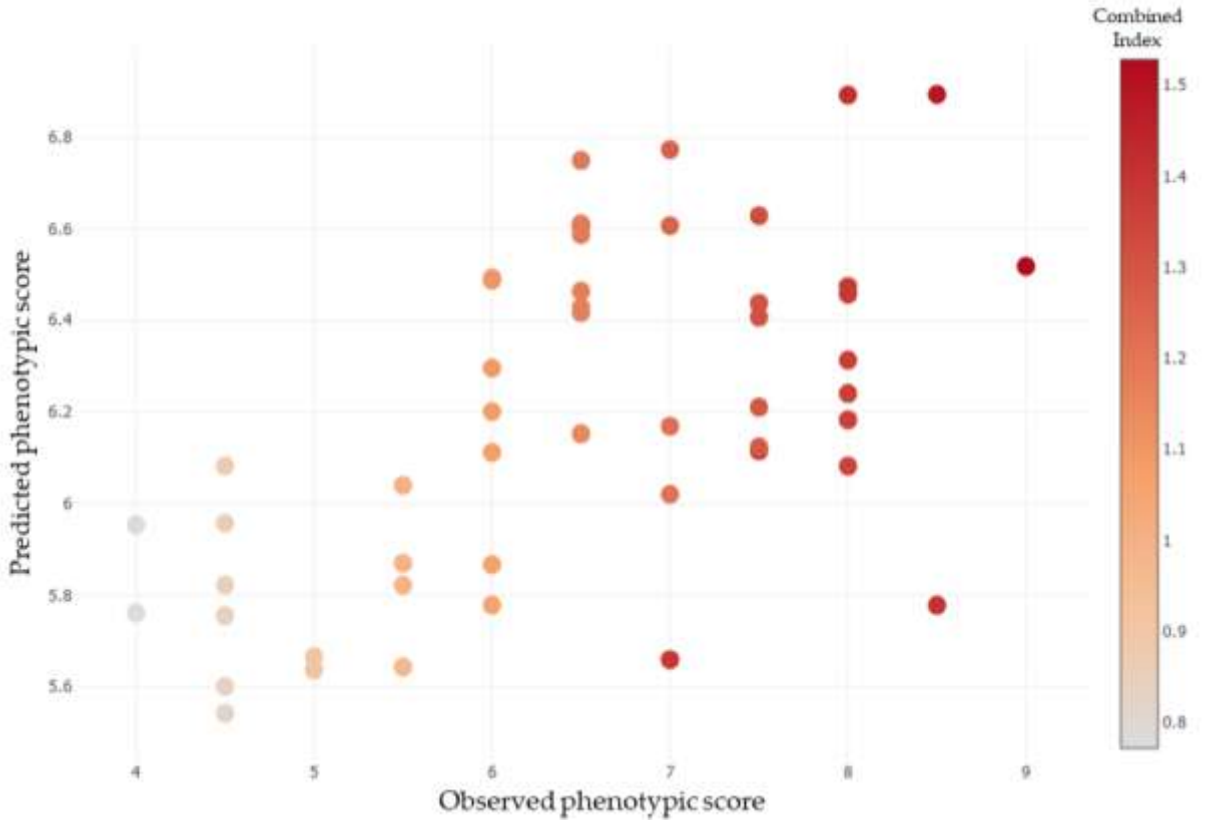


Figure 2. Scatter plot of the observed and predicted disease score values from the genomic prediction. Indicating the combined index of each individual, based on the aforementioned traits, as well as the Dragendorff score. Lower scores (lighter colors) indicate more valuable individuals based on the combined index score.

The combined index facilitates the identification of individuals combining low alkaloid content with favorable resistance profiles, while also highlighting discrepancies between observed and predicted performance.

Overall, these findings support a combined strategy of marker-assisted selection for simple traits and genomic prediction for complex traits, enabling simultaneous improvement across traits with distinct genetic architectures.

Conclusion

This thesis provides new insights into the genetic architecture of two major breeding targets in white lupin: low quinolizidine alkaloid (QA) content and anthracnose resistance. The discovery of a novel low-QA determinant on chromosome *Lalb_Chr05*, together with its additive interaction with the *pauper* locus enabled the identification of stacked allele recombinants with exceptionally low alkaloid content. These findings demonstrate the effectiveness of marker-assisted selection for traits controlled by major loci and provide valuable tools for the development of improved white lupin cultivars for food applications without de-bittering.

In contrast, anthracnose resistance exhibited a complex polygenic architecture. The limited transferability of previously published markers and the reduced predictive performance across independent populations highlight the challenges of breeding for quantitative disease resistance. Nevertheless, the improvement of genomic prediction accuracy through model retraining demonstrates the potential of integrating population-specific phenotypic and genotypic information into breeding pipelines.

Overall, the results support a combined breeding strategy using marker-assisted selection for simple traits and genomic prediction for complex traits. Future work should focus on fine-mapping and functional validation of the novel low-QA determinant, as well as evaluating the agronomic performance and ecological resilience of ultra-low alkaloid genotypes under field conditions. For anthracnose resistance, expanding training populations, increasing marker density, and integrating multi-environment phenotypic datasets will be essential to improve prediction robustness and transferability.

Together, these approaches will contribute to the development of white lupin cultivars combining improved nutritional quality with durable disease resistance.

Future prospects

Regarding MAS for white lupin genetic improvement, the molecular markers generated in this study are included in two larger marker assays within two separate European projects.

An extended list of molecular markers was provided to the EU project BELIS team (https://cordis.europa.eu/project/id/1010818789) as contribution for a legume Multi-Species Genotyping Array for Five Grain Legume Crops including 11,009 markers for white lupin.

The marker set developed in Patyi el al. (2026) will be further exploited in the EU project IPMorama (https://cordis.europa.eu/project/id/101135348), which aims to advance the applicability of genomic selection for anthracnose resistance in white lupin breeding. The SNPS from this study were combined in an extended set of 137 markers distributed along the 25 white lupin chromosomes. GS models based on this extended set are under validation with genotyping of ca 300 unknown accessions and validation with phenotyping of a subset of accessions based on the criterion of country of origin. From each subset, accessions predicted to show high, intermediate, and low resistance were selected, resulting in a total of 60 accessions currently being evaluated in field trials. The aim is to compare observed field performance with predicted resistance values to assess the accuracy and usefulness of the genomic prediction models.

Moreover, selected white lupin lines, validated in this study for improved and reproducible anthracnose resistance, together with two susceptible references, were subjected for differential gene expression profiling by transcriptome sequencing in a controlled environment, to decipher molecular background underlying their increased resistance. This research was part of the research collaboration between FiBL and PAN-IGR conducted during this PhD project, though not published yet. The climate chamber experiment, RNA isolation, and downstream analysis from count data normalization onwards were performed within the scope of this PhD.

This research, when finalized, should provide knowledge about genetic components involved in a successful response to inoculation in the early stages of the infections, highlighting gene co-expression networks established during the progression of defense response. Therefore, besides anticipated future implementation in white lupin breeding industry, presented dissertation laid the foundation for tackling basic science research in plant-pathogen interactions, aiming to compare defense responses between different anthracnose resistance donors and susceptible accessions.

References

Alkemade, J. A., Messmer, M. M., Arncken, C., Leska, A., Annicchiarico, P., Nazzicari, N., Książkiewicz, M., Voegele, R. T., Finckh, M. R., & Hohmann, P. (2021). A High-Throughput Phenotyping Tool to Identify Field-Relevant Anthracnose Resistance in White Lupin. *Plant Disease*, *105*(6), 1719–1727. https://doi.org/10.1094/PDIS-07-20-1531-RE

Alkemade, J. A., Messmer, M. M., Voegele, R. T., Finckh, M. R., & Hohmann, P. (2021). Genetic diversity of Colletotrichum lupini and its virulence on white and Andean lupin. *Scientific Reports*, *11*(1), 13547. https://doi.org/10.1038/s41598-021-92953-y

Alkemade, J. A., Nazzicari, N., Messmer, M. M., Annicchiarico, P., Ferrari, B., Voegele, R. T., Finckh, M. R., Arncken, C., & Hohmann, P. (2022). Genome-wide association study reveals white lupin candidate gene involved in anthracnose resistance. *Theoretical and Applied Genetics*, *135*(3), 1011–1024. https://doi.org/10.1007/s00122-021-04014-7

Alkemade, J. A., Patyi, A., Arncken, C., & Lazzaro, M. (2025). Lupins: A Remarkable Protein Crop Battling Anthracnose for a Greener Future. *Plant Health Cases*, *2025*, phcs20250008. https://doi.org/10.1079/planthealthcases.2025.0008

Barrera, L. L., Trujillo, M. E., Goodfellow, M., García, F. J., Hernández-Lucas, I., Dávila, G., Van Berkum, P., & Martínez-Romero, E. (1997). Biodiversity of bradyrhizobia nodulating Lupinus spp. *International Journal of Systematic Bacteriology*, *47*(4), 1086–1091. https://doi.org/10.1099/00207713-47-4-1086

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Blighe, K., & Lun, A. (2023). *PCAtools: PCAtools: Everything Principal Components Analysis. Doi:10.18129/B9.bioc.PCAtools* (Version 2.14.0) [Computer software]. https://bioconductor.org/packages/PCAtools

Book of abstract. (2025). 14th conference of the European foundation for plant pathology - visions for the future, Uppsala 2-5 June 2025. https://doi.org/10.54612/a.1pdfrti5u6

Bundesinstitut Für Risikobewertung. (2017). *Risk assessment of the occurrence of alkaloids in lupin seeds*. https://doi.org/10.17590/20170530-142504

Endelman, J. (2023). *rrBLUP: Ridge Regression and Other Kernels for Genomic Selection* (Version 4.6.3) [Computer software]. https://cran.r-project.org/web/packages/rrBLUP/index.html

Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, *201*(4), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Gondran, J., Bateman, G. L., Milford, G. F. J., Bayer, J., Beerepoot, L., Boller, B., Caligari, P. D. S., Carrasco Lopez, J. M., Crowley, J. G., Rocha, J. da, Feiler, U., Gataulina, G. G., Golovchenko, O. V., Korneichuk, N. S., Frencel, I., Jaubertie, J. P., Jeffes, M., Jordan, A. C., Jörnsgaard, B., … Szukala, J. (et al). (1999). Anthracnose of white lupin (Lupinus albus): European prospects for a future sustainable crop. In G. D. Hilleditor (Ed.), *Towards the 21st century. Proceedings of the 8th international lupin conference, Asilomar, California, USA, May 11-16 1996*. (pp. 512–519). Canterbury, International Lupin Association.

Guilengue, N., Silva, M. D. C., Talhinhas, P., Neves-Martins, J., & Loureiro, A. (2022). Subcuticular–Intracellular Hemibiotrophy of Colletotrichum lupini in Lupinus mutabilis. *Plants*, *11*(22), 3028. https://doi.org/10.3390/plants11223028

Hondelmann, W. (1984). The lupin—Ancient and modern crop plant. *Theoretical and Applied Genetics*, *68*(1–2), 1–9. https://doi.org/10.1007/BF00252301

Hufnagel, B., Marques, A., Soriano, A., Marquès, L., Divol, F., Doumas, P., Sallet, E., Mancinotti, D., Carrere, S., Marande, W., Arribat, S., Keller, J., Huneau, C., Blein, T., Aimé, D., Laguerre, M., Taylor, J., Schubert, V., Nelson, M., … Péret, B. (2020). High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nature Communications*, *11*(1), 492. https://doi.org/10.1038/s41467-019-14197-9

Hufnagel, B., Soriano, A., Taylor, J., Divol, F., Kroc, M., Sanders, H., Yeheyis, L., Nelson, M., & Péret, B. (2021). Pangenome of white lupin provides insights into the diversity of the species. *Plant Biotechnology Journal*, *19*(12), 2532–2543. https://doi.org/10.1111/pbi.13678

Huyghe, C. (1997). Field Crops Research White lupin (Lupinus albus L.). *Field Crops Research*, *53*, 147–160.

Kroc, M., Rybiński, W., Wilczura, P., Kamel, K., Kaczmarek, Z., Barzyk, P., & Święcicki, W. (2017). Quantitative and qualitative analysis of alkaloids composition in the seeds of a white lupin (Lupinus albus L.) collection. *Genetic Resources and Crop Evolution*, *64*(8), 1853–1860. https://doi.org/10.1007/s10722-016-0473-1

Książkiewicz, M., Nazzicari, N., Yang, H., Nelson, M. N., Renshaw, D., Rychel, S., Ferrari, B., Carelli, M., Tomaszewska, M., Stawiński, S., Naganowska, B., Wolko, B., & Annicchiarico, P. (2017). A high-density consensus linkage map of white lupin highlights synteny with narrow-leafed lupin and provides markers tagging key agronomic traits. *Scientific Reports*, *7*(1), 15335. https://doi.org/10.1038/s41598-017-15625-w

Lenth, R. V. (2017). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (p. 1.10.7) [Dataset]. https://doi.org/10.32614/CRAN.package.emmeans

Lenth, R. V., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Piaskowski, J., Riebl, H., & Singmann, H. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.10.7) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Lin, R., Renshaw, D., Luckett, D., Clements, J., Yan, G., Adhikari, K., Buirchell, B., Sweetingham, M., & Yang, H. (2009). Development of a sequence-specific PCR marker linked to the gene "pauper" conferring low-alkaloids in white lupin (Lupinus albus L.) for marker assisted selection. *Molecular Breeding*, *23*(1), 153–161. https://doi.org/10.1007/s11032-008-9222-2

Mancinotti, D., Czepiel, K., Taylor, J. L., Golshadi Galehshahi, H., Møller, L. A., Jensen, M. K., Motawia, M. S., Hufnagel, B., Soriano, A., Yeheyis, L., Kjaerulff, L., Péret, B., Staerk, D., Wendt, T., Nelson, M. N., Kroc, M., & Geu-Flores, F. (2023). The causal mutation leading to sweetness in modern white lupin cultivars. *Science Advances*, *9*(31), eadg8866. https://doi.org/10.1126/sciadv.adg8866

Patyi, A., Kamp, M., Arncken, C., Biazzi, E., Książkiewicz, M., Messmer, M. M., Schneider, M., Tava, A., & Lazzaro, M. (2025). Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (Lupinus albus, L.) and development of ultra-low alkaloid recombinants by stacking with the pauper allele. *BMC Plant Biology*, *25*(1), 945. https://doi.org/10.1186/s12870-025-06951-7

Patyi, A., Schneider, M., Arncken, C., Książkiewicz, M., Messmer, M. M., Schwertfirm, G., & Lazzaro, M. (2026). From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments. *Molecular Breeding*, *46*(5), 41. https://doi.org/10.1007/s11032-026-01661-w

Philippi, J., Schliephake, E., Jürgens, H.-U., Jansen, G., & Ordon, F. (2015). Feeding behavior of aphids on narrow-leafed lupin (Lupinus angustifolius) genotypes

varying in the content of quinolizidine alkaloids. *Entomologia Experimentalis et Applicata*, *156*(1), 37–51. https://doi.org/10.1111/eea.12313

Philippi, J., Schliephake, E., Jürgens, H.-U., Jansen, G., & Ordon, F. (2016). Correlation of the alkaloid content and composition of narrow-leafed lupins (Lupinus angustifolius L.) to aphid susceptibility. *Journal of Pest Science*, *89*(2), 359–373. https://doi.org/10.1007/s10340-015-0710-y

Prusinski, J. (2017). White lupin (Lupinus albus L.)—Nutritional and health values in human nutrition—A review. *Czech Journal of Food Sciences*, *35*(2), 95–105. https://doi.org/10.17221/114/2016-CJFS

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>.

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>.

Robinson, K. O., Beyene, D. A., Van Berkum, P., Knight-Mason, R., & Bhardwaj, H. L. (2000). Variability in plant-microbe interaction between Lupinus lines and Bradyrhizobium strains. *Plant Science*, *159*(2), 257–264. https://doi.org/10.1016/S0168-9452(00)00345-9

Rychel, S., & Książkiewicz, M. (2019). Development of gene-based molecular markers tagging low alkaloid pauper locus in white lupin (Lupinus albus L.). *Journal of Applied Genetics*, *60*(3–4), 269–281. https://doi.org/10.1007/s13353-019-00508-9

Schneider, M., Shrestha, A., Ballvora, A., & Léon, J. (2022). High-throughput estimation of allele frequencies using combined pooled-population sequencing and haplotype-based data processing. *Plant Methods*, *18*(1), 34. https://doi.org/10.1186/s13007-022-00852-8

Schrenk, D., Bodin, L., Chipman, J. K., del Mazo, J., Grasl-Kraupp, B., Hogstrand, C., Hoogenboom, L., Leblanc, J. C., Nebbia, C. S., Nielsen, E., Ntzani, E., Petersen, A., Sand, S., Schwerdtle, T., Vleminckx, C., Wallace, H., Alexander, J., Cottrill, B., Dusemund, B., … Bignami, M. (2019). Scientific opinion on the risks for animal and human health related to the presence of quinolizidine alkaloids in feed and food, in particular in lupins and lupin-derived products. *EFSA Journal*, *17*(11). https://doi.org/10.2903/j.efsa.2019.5860

Schrenk, D., Bodin, L., Chipman, J. K., del Mazo, J., Grasl-Kraupp, B., Hogstrand, C., Hoogenboom, L. (Ron), Leblanc, J., Nebbia, C. S., Nielsen, E., Ntzani, E., Petersen, A., Sand, S., Schwerdtle, T., Vleminckx, C., Wallace, H., Alexander, J.,

Cottrill, B., Dusemund, B., … Bignami, M. (2019). Scientific opinion on the risks for animal and human health related to the presence of quinolizidine alkaloids in feed and food, in particular in lupins and lupin-derived products. *EFSA Journal*, *17*(11), e05860. https://doi.org/10.2903/j.efsa.2019.5860

Schryvers, S., Arinzechukwu, C., Miserez, B., Eeckhout, M., & Jacxsens, L. (2023). The fate of quinolizidine alkaloids during the processing of lupins (*Lupinus* spp.) for human consumption. *Food Chemistry*, *429*, 136847. https://doi.org/10.1016/j.foodchem.2023.136847

Schwertfirm, G., Schneider, M., Haase, F., Riedel, C., Lazzaro, M., Ruge-Wehling, B., & Schweizer, G. (2024). Genome-wide association study revealed significant SNPs for anthracnose resistance, seed alkaloids and protein content in white lupin. *Theoretical and Applied Genetics*, *137*(7), 155. https://doi.org/10.1007/s00122-024-04665-2

Sreenivasaprasad, S., & Talhinhas, P. (2005). Genotypic and phenotypic diversity in Colletotrichum acutatum, a cosmopolitan pathogen causing anthracnose on a wide range of hosts. *Molecular Plant Pathology*, *6*(4), 361–378. https://doi.org/10.1111/j.1364-3703.2005.00291.x

Talhinhas, P., Baroncelli, R., & Le Floch, G. (2016). ANTHRACNOSE OF LUPINS CAUSED BY COLLETOTRICHUM LUPINI: A RECENT DISEASE AND A SUCCESSFUL WORLDWIDE PATHOGEN. *Journal of Plant Pathology*, *98*(1). https://doi.org/10.4454/JPP.V98I1.040

Thomas, G. J., & Sweetingham, M. W. (2004). Cultivar and environment influence the development of lupin anthracnose caused by Colletotrichum lupini. *Australasian Plant Pathology*, *33*(4), 571–577. https://doi.org/10.1071/AP04060

Van De Noort, M. (2017). Lupin. In *Sustainable Protein Sources* (pp. 165–183). Elsevier. https://doi.org/10.1016/B978-0-12-802778-3.00010-X

Words, I. (1982). *GENETICAL C O N T R O L OF A L K A L O I D S IN*. *31*, 357–364.

Xu, W., Zhang, Q., Yuan, W., Xu, F., Muhammad Aslam, M., Miao, R., Li, Y., Wang, Q., Li, X., Zhang, X., Zhang, K., Xia, T., & Cheng, F. (2020). The genome evolution and low-phosphorus adaptation in white lupin. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-14891-z

Publication nr. 1

Patyi, A., Kamp, M., Arncken, C., Biazzi, E., Książkiewicz, M., Messmer, M. M., Schneider, M., Tava, A., & Lazzaro, M. (2025). Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (Lupinus albus, L.) and development of ultra-low alkaloid recombinants by stacking with the pauper allele. BMC Plant Biology, 25(1), 945. https://doi.org/10.1186/s12870-025-06951-7

Journal metrics:

- Impact Factor (2024): 4.8
- 5-year journal impact factor (2024): 5.4
- MNiSW: 140

BMC Plant Biology

RESEARCH Open Access

Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele

András Patyi[1,2*], Miriam Kamp[3], Christine Arncken[1], Elisa Biazzi[4], Michał Książkiewicz[2], Monika M. Messmer[1], Michael Schneider[1], Aldo Tava[4] and Mariateresa Lazzaro[1]

Abstract

Background White lupin (*Lupinus albus*, L.) is a grain legume cultivated for its high agronomic and nutritional potential, despite the accumulation of bitter and potentially harmful to health quinolizidine alkaloids (QA) in the grain. Modern sweet (i.e. low alkaloid content) varieties exist, which are exploiting different recessive mutations responsible for the desired low QA chemotype. The most widely used QA-reducing determinant, *pauper*, has been recently identified enabling marker-assisted selection, but unstable QA content across growing seasons and environments remain a challenge in white lupin cultivation.

Results Through Bulked Segregant Analysis of an F_2 population, segregating for two different low QA conferring loci, we identified a novel QTL spanning a 1 Mbp region on chromosome 5, a novel source of sweetness apart from *pauper*. We present marker-trait associations for the new locus tagging low QA content in white lupin within this QTL, described in F_2 generation and validated in F_3. Together with genotyping of sweetness tagging *pauper* sweet individuals, we identified, in F_3, 50 stacked allele recombinants where the low QA chemotype is further reduced. These individuals exhibit an exceptionally low total alkaloid content (22.8 ± 10.4 ppm), even when compared to genotypes known to carry the allele associated to the most drastic reduction in QAs, *pauper* (171.7 ± 18.5 ppm).

Conclusion The discovery of this novel locus and the development of associated PACE markers, potentially applicable for marker-assisted selection together with *pauper* (especially after farther validation in a larger panel of accessions), can enhance the improvement of white lupin through the development of new varieties with very low

*Correspondence:
András Patyi
andras.patyi@fibl.org

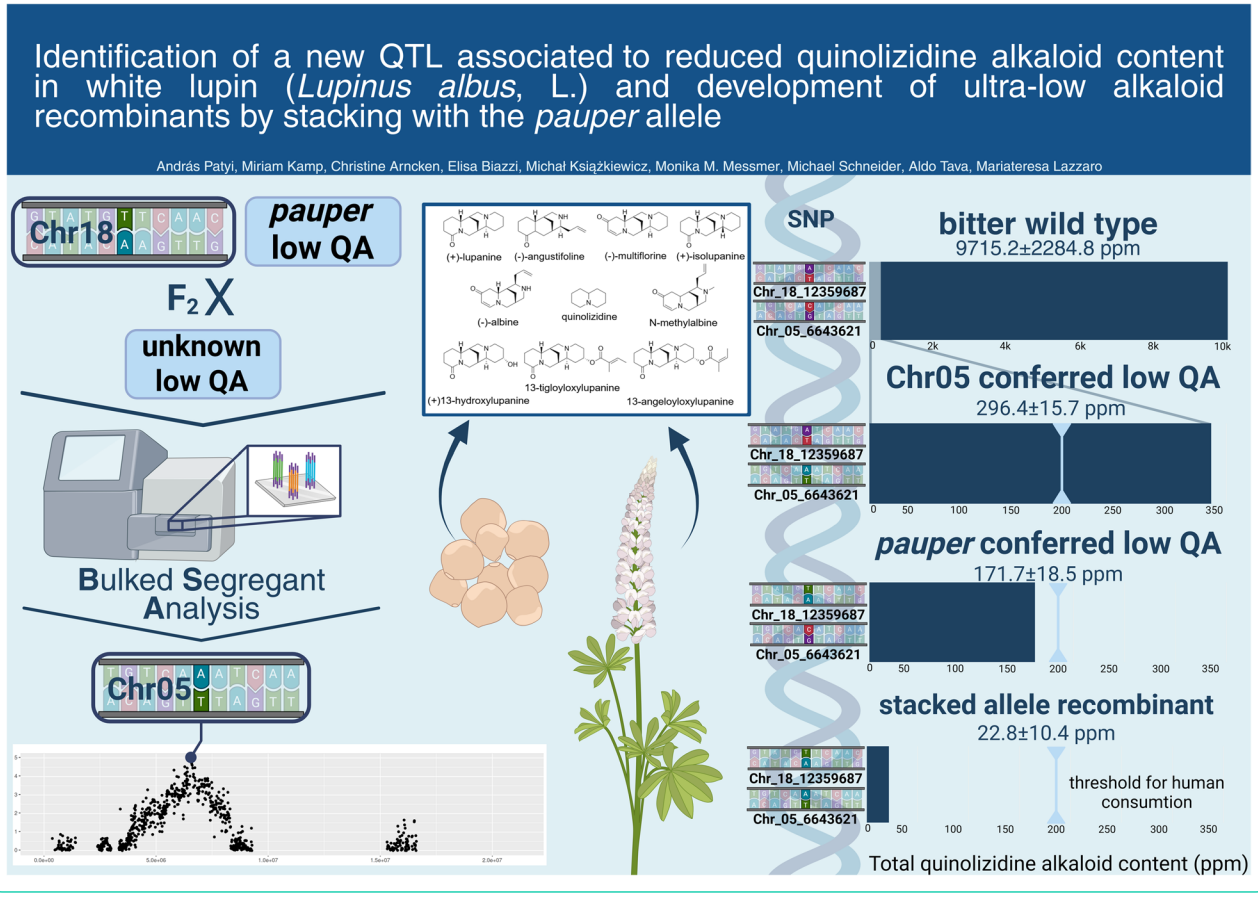
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

and stable alkaloid content based on stacked allele recombinants. This can help to increase the cultivation of this useful yet underutilised crop and its use for human nutrition.

Keywords White lupin, Bulked Segregant analysis (BSA), Quantitative trait locus (QTL), Quinolizidine alkaloids (QA), PCR allelic competitive extension (PACE), Plant breeding

Graphical abstract



Background

White lupin (*Lupinus albus*, L.) is a grain legume of Mediterranean origin produced for its agronomic value and highly nutritious seeds. From the nutritional perspective, white lupin grain is appreciated for feed and food for its high protein (33–47% depending on genotype and location) and excellent fibre content [1]. Its desirable fatty acid composition ratio of omega-3 to omega-6, non-starch carbohydrate-, oligosaccharide- and antioxidant-content makes the consumption of white lupin seed a great contribution to a healthy diet [2]. Conversely, white lupin seed contains quinolizidine alkaloids (QAs), toxic (primarily due to their ability to interfere with the nervous system) and bitter compounds which set a major challenge for food producers and processors. In the closely-related blue lupin (*Lupinus angustifolius*, L.), the major site of production of QAs is the aerial epidermis, later transported to seeds during fruit development [3]. Seeds of wild forms of white lupin contain up to 127,000 ppm (mg/kg dry weight) QAs which is reduced in bitter white lupin cultivars to 2,000–4,000 ppm [4]. Historically, the grains were soaked in running water to remove the toxic alkaloids before consumption. Improved methods exist for debittering, commonly using water as a solvent, however these processes remain water-use intensive, laborious and producing toxic waste [5], constraining the use and economic viability of white lupin in the value chain, for feed and, especially, food.

Lupins contain a mixture of different alkaloids, not all of which are sufficiently known for their toxicity. Recently, Schreiber et al. [6] evaluated the toxicity of five common lupin alkaloids (sparteine, lupinine, lupanine, 13-hydroxylupanine, angustifoline) reporting that these QAs individually and as a mixture are neither cyto-, nor genotoxic in mammalian cells. Concerning their neurotoxicity, the effect of sparteine, presumably

the most toxic lupin alkaloid, has been studied best. The available risk assessment is therefore based on the knowledge of sparteine for all alkaloids present in lupins. Safety control is based on the total quinolizidine alkaloid content [7]. In Switzerland, for example, there are no legal limits for alkaloids in food or feed. The principle of self-regulation applies, which means that companies must ensure that only safe food is placed on the market [8]. The German Federal Institute for Risk Assessment (BfR) has established in 2017 the following guideline values to protect human and animal health: <200 ppm and < 500 ppm (dry weight) respectively for food and feed, relative to the final product intended for consumption [9]. Additionally, the import and export standards of the industry organisation Pulse Australia (Australia being the world's largest lupin producer) set in 2001 a maximum alkaloid content of 200 ppm [7]. The European Commission has recently recommended monitoring of the following quinolizidine alkaloids in lupins and lupin-derived food: albine, anagyrin, angustifoline, lupanine, isolupanine, multiflorine, 13α-hydroxylupanine, lupinine and sparteine [10].

Modern sweet (i.e. bitterness lacking) varieties exist with the alkaloid content ranging between 100 and 800 ppm of seed dry weight [11]. Low QA cultivars were a major achievement of the formal breeding of white lupin, started in the early 20th century with the activity of Reinhold von Sengbusch [12]. He selected against common wild type traits, amongst many others for high QA content, describing the drastic decrease (to 200–500 ppm) in QA content due to the recessive allele *pauper*, present in many modern white lupin cultivars [13]. Apart from *pauper*, other loci exist (e.g. *exiguus*, *nutricius*, *mitis*, *reductus*) representing possible recessive mutations for low QA content, described to reduce QA content to ≤1000 ppm [11]. These mutations were described based on crossings between different white lupin individuals and observing the segregation pattern in the chemotype, thus not much is known about the genes involved or transcriptomic regulation of low QA content in white lupins, except for *pauper*. Recently, the mutation causing the *pauper*-type sweetness in white lupin was identified [14]. Reportedly the presence of a single SNP (*Lalb_Chr18_12359687*) confers a greatly reduced total QA content and a distinct QA profile. This mutation is located in the coding sequence of a gene encoding an acetyltransferase, and is responsible for changes in the early biosynthetic steps of QAs, resulting in the low QA chemotype. However, phenotypically sweet variants exist, which do not carry the *pauper* mutation, confirming non-*pauper* sweetness types, earlier described in the 20th century [11]. One example was the cultivar 'Dieta', in which the genetic determinant causing low QA levels is unknown [14].

Despite the magnitude of the effect of the gene associated with *pauper*, grains harvested from reportedly sweet *pauper* accessions have been observed to exceed the advised threshold for consumption (FiBL, unpublished data). Relatively high and unstable QA content in the harvested grain remains a threat for white lupin producers and processers, creating a demand for even lower QA varieties with stable sweetness across growing seasons regardless of the environmental context. This could be achieved through gene stacking (pyramiding)– combining independent loci conferring low QA content in white lupin, which could be further aided with the availability of molecular markers tagging these loci.

The present study aimed to map an unknown source of sweetness in white lupin by Bulked Segregant Analysis (BSA) in an F_2 population resulting from the crossing of the *pauper* sweet cultivar 'Frieda' and the cultivar 'Dieta' carrying the unknown sweetness determinant. Subsequently, through PACE marker development and validation in a diversity panel, we aimed to verify the effect of the individual loci conferring the low QA level phenotype and to assess their combined effect on QA concentration and profile in the F_2 and the derived F_3 population.

Methods

Plant material

Crossing and F_2 population

Two commercial cultivars, 'Frieda' (♂, *pauper* sweetness) and 'Dieta' (♀, unknown sweetness), were crossed in 2021. Seeds were obtained from Delley Samen und Pflanzen AG (Delley-Portalban, Switzerland) for 'Frieda' and from Soya UK Ltd. (Southampton, United Kingdom) for 'Dieta'. The five harvested seeds from the crossing were propagated in 2022 under protective nets to prevent cross-pollination and seeds of the F_1 plants were harvested in a bulk.

The F_2 plants (own breeding material) were grown in 2023 in the FiBL breeding nursery located in Leibstadt, Switzerland (47°35'26.4"N 8°10'56.0"E), organized in a 10-row plot with 14 $F_{1:2}$ seeds sown per row, grown under protective nets. Young leaf tissue (ca. 100 mg) was collected and frozen from the 112 surviving individuals of the F_2 population (Supplementary Table S1).

F_3 population

Thirteen F_2 plants, all heterozygous at the *pauper* locus were selected for the F_3 population (own breeding material) and their seeds were grown in 2024 on a field in Full-Reuenthal, Switzerland (47°35'59.7"N 8°12'13.6"E), sown and processed in an identical manner as in previous years. In total, 329 individuals represented the F_3 population (Supplementary Table S3). In addition, 10 individuals each of 'Frieda' and 'Dieta' were grown in the same field and harvested as single plants.

Validation panel

A diverse panel of 34 individuals was created for the molecular marker validation, consisting of 1 to 7 representatives for 1 breeding line, 16 cultivars, 3 landraces and 1 wild type in addition to the 39 accessions of the white lupin pangenome [15] (Supplementary Table S5).

Phenotyping

The alkaloid content of F_2 and F_3 was estimated first on the field using Dragendorff paper test (Supplementary Table S1, Supplementary Table S3), a filter paper soaked in iodine/potassium iodide solution (Dragendorff solution), then dried at 40 °C [16]. Individual QA content was estimated by transferring plant sap from the petioles onto the Dragendorff paper during onset of flowering (BBCH 63–65). A visible reaction of the Dragendorff paper turning a reddish-brown colour was described to occur for QA contents of ≥ 500 ppm [17]. The colour reaction was scored from 1 to 4, 1 meaning a purely white stain, the lowest alkaloid content. A score of 2 was assigned to stains, which were mostly white with the red encircling, score 3 described red encircled pinkish stains with a light centre and lastly, a score of 4 was given to samples with an intense red stain without discolouration in the centre (Fig. 1).

Additionally, bitterness in the F_2 population was phenotyped by scoring the taste of one flower per plant in a range from 1 (absence of bitterness) to 9 (extremely bitter). For a more accurate phenotypic score in the F_2, we combined the Dragendorff phenotypic score multiplied by two to that of the bitterness tasting score. In this combined phenotypic score, the phenotypic value ranged from 3 to 15, 15 being the most bitter.

Chemotyping of F_3 individual seeds

Approximately 1 g of seeds (3–5 seeds) of a curated panel of F_3 (Supplementary Table S6) and parental individuals were milled for analytical QA quantification. QA extraction and characterization were performed according to modified versions of the protocol described in Wink [18], Resta et al. [19] and Boschin et al. [20]. Briefly, 100 mg of seed flour was suspended in 1.2 mL of 0.1 N HCl, with sparteine (CAS 90-39-1; Extrasynthese, France) added as internal standard in an appropriate concentration, and stirred at room temperature overnight. The mixture was centrifuged at 8000 g for 45 min at 4 °C, the supernatant was collected, and the solid was washed twice with 0.8 mL of 0.1 N HCl. The gathered extracts were alkalinized with 5% NH_4OH to pH 10 – 11 and then applied onto an Extrelut NT 3 column (Supelco, Merck, Darmstadt, Germany). The alkaloids were eluted with CH_2Cl_2 (3 × 5 mL), and the solvent was evaporated under vacuum. The residue was resuspended in an appropriate volume of dichloromethane and analyzed

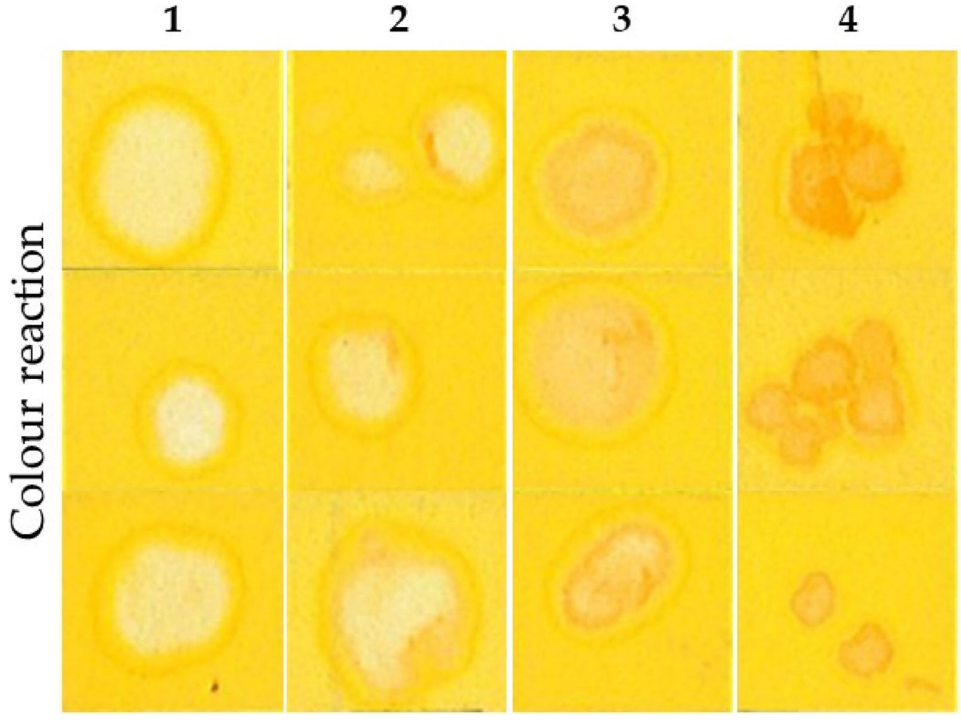


Fig. 1 Examples of the colour-reaction of Dragendorff paper according to the predefined scoring scale from 1 (very sweet) to 4 (very bitter). 1 - purely white stain; 2 - mostly white with the red encircling; 3 - red encircled pinkish stains with a light centre 4 - intense red stain without discolouration in the centre

by gas chromatography. Each sample was independently extracted, then analyzed in duplicate.

Total and individual QA content was determined quantitatively using the internal standard methods by gas chromatography/Flame Ionization Detector (Perkin-Elmer Clarus 500 GC-FID), equipped with a capillary column Elite-5 MS (DB-5, 30 m × 0.32 mm × 0.25 µm; Perkin-Elmer; Milan, Italy). The oven temperature ramp was held at 90 °C for 2 min, increased to 300 °C at 7 °C/min and held at 300 °C for 10 min. Helium was used as the carrier gas, and the flow rate was set at 2 mL/min. The temperature of the injector was set at 300 °C, and the injection volume was 1 µL; the temperatures of the FID was set at 320 °C. The absence of sparteine in the samples was demonstrated by performing GC/FID analyses without an internal standard (choosing one sample per recombinant).

The response factor of GC/FID was calculated using the ratio between the response of the internal standard (sparteine) and the response of the analyte standard lupanine. The regression coefficient between the analyte concentration and detector response was $R^2 = 0.991$.

Qualitative identification of QAs was performed by gas chromatography/mass spectrometry; GC/MS analyses were carried out using a Perkin Elmer Clarus 500 GC equipped with a Clarus 500 mass spectrometer using the same capillary column and chromatographic conditions as for the GC/FID analyses (above mentioned). Mass spectra were acquired over a range of 40–400 atomic mass units (amu) at 1 scan/sec with ionizing electron energy of 70 eV and ion source at 230 °C. The transfer line was set at 300 °C, while the carrier gas was helium at 1.0 mL/min. The QAs were identified by determination of their elution times published mass spectra [18–20], as well as a peak-matching library search [21]. The standard GC condition described above allowed a practical measurable sensitivity of 10 ng/µl per injection, i.e. a detection limit of 1 µg of lupanine g^{-1} of lupin seed flour (1 ppm).

Pooling for bulked segregant analysis

Two bulks consisting of DNA (isolated with a modified CTAB method) from eight plants of the sweet and bitter phenotypic extremes of the 'Dieta' x 'Frieda' F_2 experimental population were curated for sequencing. Homozygous individuals for the *pauper* locus were excluded from this subset to discriminate sweet individuals where the low QA content is conferred by *pauper*. Thus, the F_2 bulks consisted of individuals which were heterozygous or wild type at the *pauper* locus, suggesting that the low QA phenotype is conferred by a determinant other than *pauper*. The combined phenotypic score [3–15] of the sweet bulk was between 7 and 8.5, while that of the bitter one was 13–15. Additionally, bulks of the parental cultivars (from seedlings grown for this purpose from the same seed source used for the crossing) consisting of eight plants each were sent for sequencing.

Sequencing-data analysis

Whole genome sequencing was performed at Novogene UK (Cambridge, UK) with a NovaSeq X Plus Series, sequencing depth of 15x. Paired-end 150 bp long reads were generated. The raw sequence data was processed and analysed in a modified version of the procedure published by Schneider et al. 2022 [22]. Quality checks were performed with the *fastqc* tool [23]. Raw reads were aligned to the reference genome using *bwa mem* [24]. Subsequently, duplicated reads were removed using the markdub-function of *sambamba* [25] and sorted by their mapped position. Reads that had been mapped to multiple locations (e.g. secondary alignments) were discarded. The variant calling was conducted using *bcftools* only on SNPs that had been previously reported [15] to improve the confidence despite the low sample number and read depth. Furthermore, reads with a quality threshold below 25 for the whole read, and 30 per SNV base call were omitted, as well as SNPs with read depths below 8. The allelic depth (AD) was extracted and heterozygous SNPs in the parental bulks were omitted from further analysis. The remaining SNPs were assigned to annotated genes based on their position in the gene or in close proximity to it, as described by Schneider et al. 2022 [22]. For the resulting 23,396 gene-anchored haplotypes, the allele frequencies of 'Dieta' and 'Frieda' were calculated for the sweet and for the bitter bulk. The frequencies were aggregated with the read depth to count values, as required for *BayPass* [26], which is a Bayesian approach to identify i.e. selective sweeps between the sweet and bitter populations. *BayPass* was conducted using the effective population size of 200 to reflect the size of the 'Dieta' × 'Frieda' F_2 population. From the p-values of each haplotype and their mapped positions (Supplementary Table S8), Manhattan plots were generated, with a threshold of $-\log_{10}(\text{p-value}) > 4.5$.

Genotyping

PACE primers

Primer triplets (Table 1) for 3 selected SNPs in the newly identified QTL and for the SNP associated to *pauper* were designed by 3CR Bioscience (Essex, UK) and oligonucleotides were acquired from IDT (Coralville IA, USA). For all four loci, the primers were designed based on the sequence of the sense strand.

PACE assay

For the genotyping assays, we used the PACE master mix (2x concentration, standard ROX level) provided by 3CR Bioscience (Essex, UK). Each individual reaction

Table 1 PACE primers used in the study. Target SNP locus on the chromosome, the oligonucleotide sequence for the two forward (F1 and F2) tagging the two possible alleles at the respective locus (underlined), including the sequence of the X- and Y- Tails and the reverse (R) primers for each triplet, and melting temperatures

Locus	Primer	Sequence (5′–3′)	Tm (°C)
Lalb_Chr05_6643621	F1	GAAGGTGACCAAGTTCATGCTAGACTATGTCATGTTTTTAAACATGTCA**C**	64.2
	F2	GAAGGTCGGAGTCAACGGATTGAGACTATGTCATGTTTTTAAACATGTCA**A**	65.4
	R	GGCTTGAATAGGGTCCATGTGTTTAATAT	57.1
Lalb_Chr05_6643864	F1	GAAGGTGACCAAGTTCATGCTAATCTATTATTGTCCTATTAAAACATTACGA**G**	62.8
	F2	GAAGGTCGGAGTCAACGGATTAATCTATTATTGTCCTATTAAAACATTACGA**C**	63.3
	R	TATATTTTTTAGCTTAGGATGTTTTGGATT	52.3
Lalb_Chr05_6645052	F1	GAAGGTGACCAAGTTCATGCTCGATAAGAGAGTGGGAGGACCA**A**	67.7
	F2	GAAGGTCGGAGTCAACGGATTGATAAGAGAGTGGGAGGACCA**C**	67.8
	R	CGTGGCATGGACCCCAGTTATATAA	58.5
Lalb_Chr18_12359687 (pauper)	F1	GAAGGTGACCAAGTTCATGCTAAATGCTATCAGGATAGGGTCTATG**T**	65.3
	F2	GAAGGTCGGAGTCAACGGATTAAATGCTATCAGGATAGGGTCTATG**A**	65.7
	R	CTTCCATTGTTCTTCCTCTATCTRCACTT	57.3

consisted of 5 µl of the PACE master mix, 0.14 µl of the primer pool and 5 µl of ELGA water (VWS Ltd.; High Wycombe, UK) or DNA template, isolated from leaf samples transferred onto Qiagen FTA cards (QIAGEN GmbH; Hilden, Germany), processed according to the manufacturer`s instructions. The cycling conditions were set according to the manufacturer`s instructions of the analogously functioning Kompetitive Allele Specific PCR (KASP) assay (LGC Group; Teddington, UK) and ran on the CFX Opus 384 Real-Time PCR System (Bio-Rad Laboratories, Inc.; Hercules CA, USA).

All 112 F_2 individuals were genotyped in 2023 using a newly designed PACE marker tagging the most important SNP associated to *pauper* [14], *Lalb_Chr18_12359687* and only retrospectively genotyped in 2024 upon the availability of the PACE marker tagging the Dieta-associated sweetness. Samples from the F_3 population were genotyped with both markers in 2024.

Data analysis

All data except for the sequencing data was analyzed and visualized in R version 4.4.2 [27].

The phenotypic data (Dragendorff score) was analyzed in the F_2 and F_3 experimental populations using Linear Mixed Models (LMM). The allelic combinations at loci *Lalb_Chr18_12359687* and *Lalb_Chr05_6643621* were included as fixed effect, and the row the specific plant belonged to on the field as a random effect. The estimated marginal means from the above-mentioned linear mixed model were calculated to test the statistical difference ($p < 0.05$) using emmeans_1.10.7 [28] and were used to assess the association between the allelic states and the phenotypic value in F_2 and F_3.

Similarly, the total alkaloid content (sum of all QAs present in a sample) and the ratio between the concentrations of 13α-hydroxylupanine and lupanine per sweet variant were analysed LMM using lme4_1.4–36 [29]. For both response variables we included the allelic combinations as fixed, and the biological replicate (i.e. set of samples belonging to allelic combinations of the two loci of interest) as random factor. The estimated marginal means from the above-mentioned linear mixed model were calculated to test the statistical difference ($p < 0.05$) using emmeans_1.10.7 [28] and were used for the following downstream analysis.

Results

The phenotypic segregation in the F_2 generation confirms another determinant for low QA content

With *pauper* known to segregate in the 'Dieta' x 'Frieda' F_2 population by the genotyping at the causal mutation SNP *Lalb_Chr18_12359687*, the simplest hypothesis of 1: 3 (sweet: bitter) phenotypic segregation ratio could be rejected. Instead, the phenotypic segregation pattern was 7: 9 (sweet: bitter), which coupled with observation of the genotypic segregation 1: 2: 1 at the *pauper* locus suggests the segregation of at least one additional low QA conferring gene aside from *pauper* (Table 2).

Assuming segregation of two independent recessive genes, the phenotypic scores are expected to segregate in a ratio of 7: 9 (sweet: bitter) in QA content (following Mendelian inheritance), which hypothesis could not be rejected (Table 2), indicating two independent recessive loci conferring low QA content with no segregation distortion in the F_2 population used for the BSA. The assumption of genotypic segregation of 1: 2: 1 could not be rejected for either loci, further supporting the independence of these determinants.

Identification of a QTL associated to low QA content through BSA

After the confirmation of the presence of a determinant independent of *pauper* for low QA content in the

Table 2 Segregation ratios and χ2 statistics in the phenotype (Dragendorff score) and at the *pauper*, and Dieta-associated locus for the entire F_2 population

Segregating population		Expected ratio	Observed ratio	df	χ2	*p*-value	
Phenotypic segregation							
'Dieta'x 'Frieda'		1:3	41:69	1	8.8364	0.00295	**
		7:9	41:69	1	1.8753	0.1709	
Genotypic segregation							
'Dieta' x 'Frieda'	*pauper* segregation	1:2:1	31:47:34	2	3.054	0.2172	
	Dieta-associated segregation	1:2:1	22:63:27	2	2.1964	0.3335	

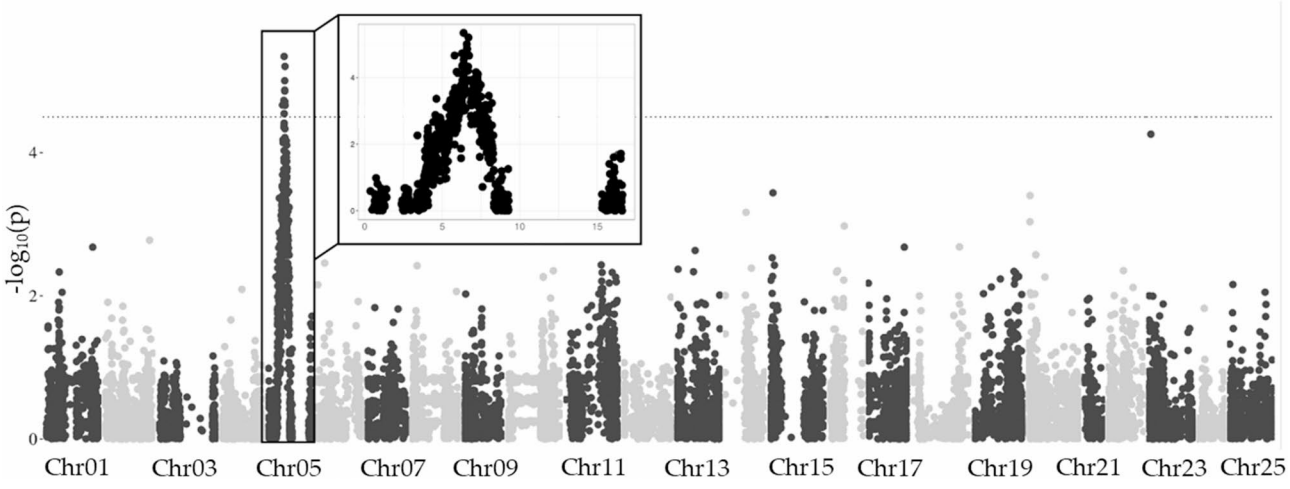


Fig. 2 Manhattan plot of -log10(p-value) (y-axis, threshold set to -log10(p-value) = 4.5), for differences in allele frequencies of haplotypes between two bulks of the F_2 population representing high and low QA levels. The x-axis indicates the position of the haplotypes in each of the 25 chromosomes. A zoomed-in plot is included for chromosome 5 for detail

segregating F_2 population, we identified the genomic region responsible for this unidentified determinant of low QA content through BSA. From the raw reads ranging between 97 and 117 million per bulk (read length of 150 bp), 47–57 million remained after quality filtering, mapping and removal of redundancies (Supplementary Table S7).

There were 1.55 million polymorphic SNPs (1.76 million before q40 filtering) from the lupin browser that were polymorphic between the parents, reported by *bcftools*, from which 514,982 remained after filtering out monomorphic, low quality and low depth variants. Following, 23,396 gene-anchored haplotypes with an average of 21 SNPs and an average read depth of 302 per haplotype were constructed. The selective sweeps analysis resulted in a clear peak on chromosome 5 (Fig. 2) that allows to differentiate the 'sweet' from the 'bitter' bulk.

Candidate gene selection and marker development

Ten genes were associated with haplotypes showing significant ($-\log_{10}$(p-value) > 4.5) differences in allele frequencies between sweet and bitter bulks within the entire QTL region spanning from 5,799,140 bp to 6,737,917 bp on chromosome 5 (Supplementary Table S9). Although none of them could be directly linked to the biosynthetic pathway, the gene *Lalb_Chr05g0222381* encoding a "Putative aminoacyltransferase" (a function similar to the gene responsible for the *pauper*-conferred low QA content) was chosen for molecular marker development. We investigated all individual SNPs within as well as 1,000 bp up- and downstream of this gene.

There are 54 polymorphic loci between 'Dieta' and 'Frieda' in this region. To determine whether these are also polymorphic within other accessions as well, we collected the allelic states for both loci of all 39 sequenced accessions from the white lupin genome browser (Supplementary Table S5). Out of these 54 SNPs, only 9 are highly polymorphic within the 39 sequenced accessions [15]. Furthermore, only 3 loci are consistent with the assumed alkaloid status of these accessions, *Lalb_Chr05_6643621*, *Lalb_Chr05_6643864*, *Lalb_Chr05_6645052*. These three were tested on the validation panel of 34 accessions to further confirm their association to alkaloid status (Supplementary Table S5), in comparison to the marker-trait association of the marker tagging *pauper*. At the locus *Lalb_Chr05_6643864*, an old cultivar, Nährquell (described to carry the sweetness mutation *nutricius*) carried the allele

associated to the Dieta-type sweetness, and the wild-type allele at the other two selected loci in the region thus this locus was not chosen for further analysis. Similarly, at the locus *Lalb_Chr05_6645052*, when comparing two different phenotypically sweet individuals of the cultivar 'Start' (described to carry the sweetness mutation *exiguus*) lead to inconsistent results, thus we selected only the SNP *Lalb_Chr05_6643621* for further genotyping of the F_3 segregating population. Interestingly, French cultivars which belong to the panel of resequenced accessions, namely 'Magnus', 'Orus', 'Luxe' and 'Clovis' carried the alleles associated to sweetness both for the *pauper* locus and the newly described one on chromosome 5.

Marker-trait associations reveal low QA stacked allele recombinants

The marker-trait association between the newly described *Lalb_Chr05_6643621* SNP, the *pauper* causal mutation SNP *Lalb_Chr18_12359687* and the phenotypic score (Dragendorff score) was confirmed for the individuals in the F_2 and F_3 populations (Supplementary Figure S1-S3), in comparison to their parental lines. In F_2, we were able to statistically differentiate between individuals carrying the sweet genotype associated to *pauper*, and discriminate for bitter individuals based on the Dragendorff score (Supplementary Table S2). However, in F_3, only *pauper* sweet individuals, and bitter individuals were clearly statistically grouped together based on the Dragendorff score only (Supplementary Table S2, Supplementary Table S4). to obtain more nuanced observations on the quantitative value of alkaloids, we performed GC/MS and GC/FID analytical QA quantification in a subset of the F3 population representing different combinations of allelic states at the pauper causal mutation SNP on chromosome 18 and the locus Lalb_Chr05_6643621 on chromosome 5, used for the following downstream analysis. This subset included the two parental lines and the most important allelic combinations at both loci (Table 3) and the corresponding estimated marginal means values from the post-hoc analysis after model fitting.

Comparing the concentrations of the total alkaloid content (Fig. 3), we observed a decrease in the estimated marginal means of the total alkaloid content in all sweet chemotypes. The F_3 plants carrying the allele *pauper* as sole contributor to the low QA phenotype (PAU) were grouped together with 'Frieda' (FRA), the parent conferring the *pauper*-type sweetness (Fig. 3). Similarly, the F_3 progeny plants in which the low QA phenotype was conferred by the parent 'Dieta' (DSW), were grouped together with the parental genotype (DIA), as well as the F_3 individuals heterozygous for *pauper* and homozygous sweet at *Lalb_Chr05_6643621* (HET), though with higher average QA content compared to the plants that are sweet because of *pauper* (PAU). The stacked allele recombinants (STA) have a very low alkaloid content (range raw data 11–62.3 ppm), lower compared to either parent or progeny, indicating a clear additive effect of stacking two low QA conferring determinants. Stacked allele recombinants (STA) represented the most drastic reduction to 12% and 8% of the total average alkaloid content of the parental cultivars, 'Dieta' and 'Frieda', respectively. The total estimated marginal mean for the alkaloid content was 22.8 ± 10.4 ppm dry weight in seeds of the stacked allele recombinants, close to 10 times lower than the suggested threshold for human consumption.

QA profiles in the F_3 population highlights the stacked allele effect

Based on the presence of specific QAs in the different allelic combinations studied, we observed distinct QA profiles associated to different low-QA determinants and their combination. In the wild-type, bitter phenotype (BIT) 14 QAs were identified (Fig. 4).

Table 3 Subset of F_3 individuals analyzed with GC/MS and GC/FID, their allelic state, number of samples (n), estimated marginal means of total Quinolizidine alkaloid (QA) content (in ppm) indicating statistical grouping, and standard error (SE). BIT– bitter wild type progeny; FRA– 'Frieda' parent, homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DIA– 'Dieta' parent, wild type at *pauper*, homozygous sweet at Lalb_Chr05_6643621; PAU– homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DSW– wild type at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; HET– heterozygous at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; STA– homozygous sweet at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*

Code	Allele				*n*	QA content (ppm)	SE (ppm)
	Lalb_Chr18_12359687 (pauper)		Lalb_Chr05_6643621				
BIT (bitter progeny)	AA	wild type	CC	wild type	4	9715.2	2284.8
DSW (Dieta sweet progeny)	AA	wild type	AA	homozygous sweet	10	296.4[c]	15.7
HET	TA	heterozygous	AA	homozygous sweet	10	267.1[c]	15.7
DIA ('Dieta' parent)	AA	wild type	AA	homozygous sweet	10	257.4[c]	15.7
PAU (pauper sweet progeny)	TT	homozygous sweet	CC	wild type	7	171.7[b]	18.5
FRA ('Frieda' parent)	TT	homozygous sweet	CC	wild type	10	176.2[b]	15.7
STA (stacked allele recombinants)	TT	homozygous sweet	AA	homozygous sweet	23	22.8[a]	10.4

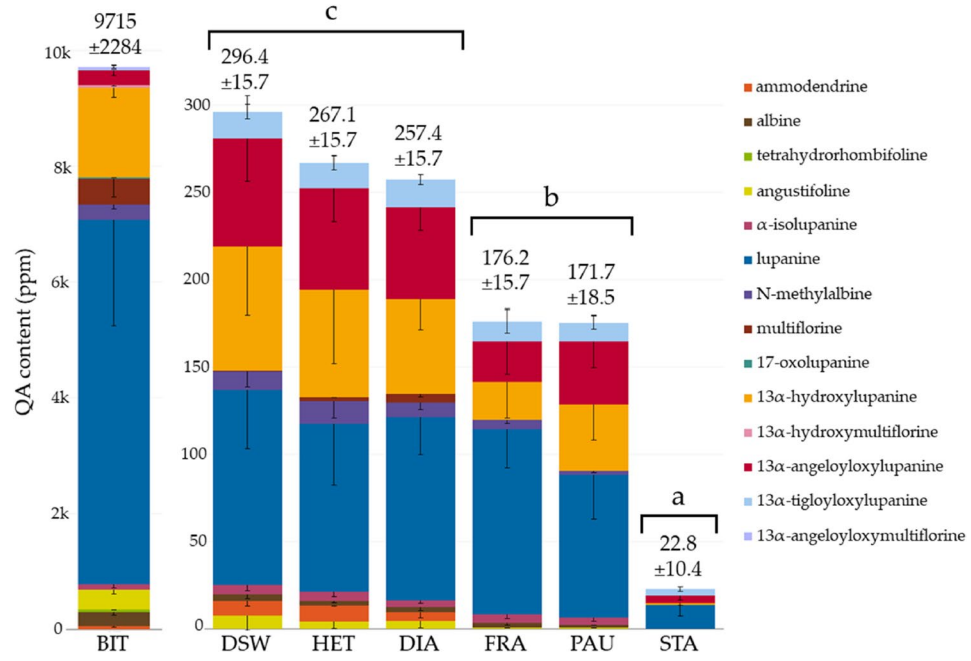


Fig. 3 Individual QA content and composition, and total QA content (estimated marginal mean value in ppm) of different recombinants in the F_3 population compared to the parental genotypes. Values with the same letter are not significantly different at 0.05 confidence level. BIT– bitter wild type progeny; FRA– 'Frieda' parent, homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DIA– 'Dieta' parent, wild type at *pauper*, homozygous sweet at Lalb_Chr05_6643621; PAU– homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DSW– wild type at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; HET– heterozygous at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; STA– homozygous sweet at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*

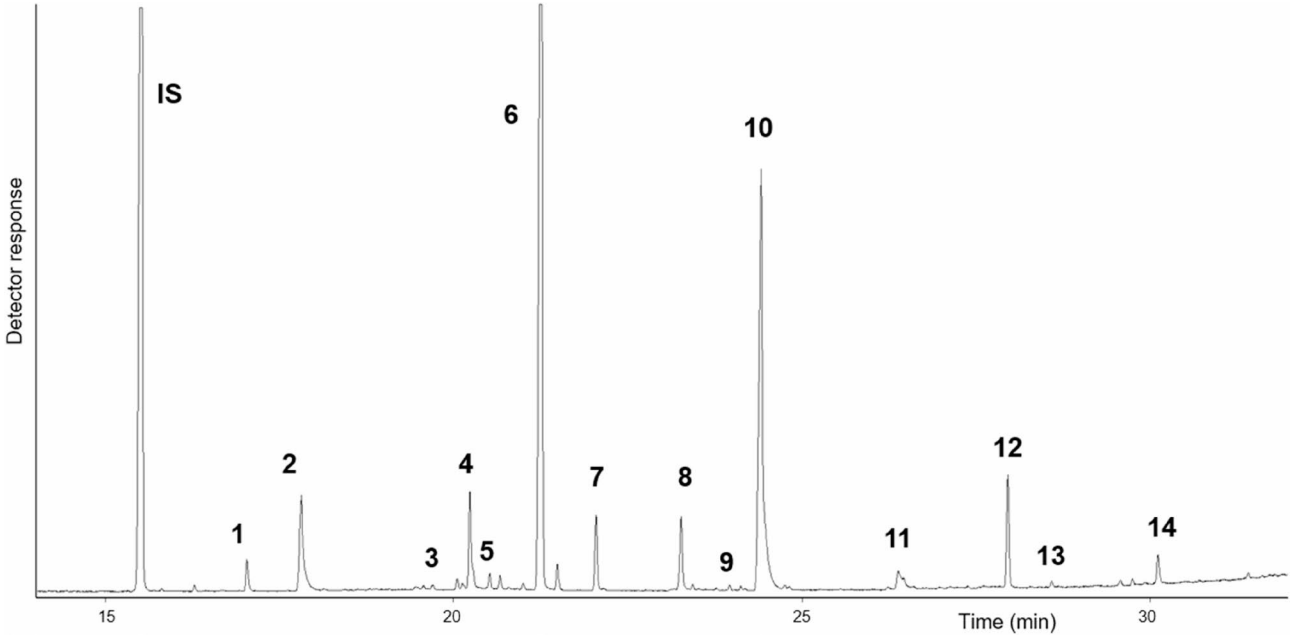


Fig. 4 Gas chromatogram of alkaloids from *L. albus* seeds. IS - internal standard (spartein); 1 -ammodendrine; 2 - albine; 3 - tetrahydrorhombifoline; 4 - angustifoline; 5 - α-isolupanine; 6 - lupanine; 7 - N-methylalbine; 8 - multiflorine; 9–17-oxolupanine; 10–13α-hydroxylupanine; 11–13α-hydroxymultiflorine; 12–13α-angeloyloxylupanine; 13–13α-tigloyloxylupanine; 14 – 13α-angeloyloxymultiflorine

Lupanine, multiflorine, 13α-hydroxylupanine, 13α-angeloyloxylupanine and 13α-tigloyloxylupanine were the most commonly occurring QAs in the bitter chemotype. Both in the individuals carrying only *pauper* (PAU) or only the 'Dieta'-associated (DIA, DSW) sweetness in a homozygous state, as well as in the line heterozygous at *pauper* and homozygous sweet at *Lalb_Chr05_6643621* (HET), there is a consistent lack of specific QAs (Fig. 5): tetrahydrorhombifoline, 17α-oxolupanine, 13α-hydroxymultiflorine and 13α-angeloyloxymultiflorine. Furthermore, the QA profile in 'Frieda' (FRA) and associated progeny (PAU) was even more reduced, lacking ammodendrine and multiflorine, while angustifoline was present only in trace amount. Stacked allele recombinants (STA) show the most drastically reduced QA profile with only lupanine and two of its esterified derivates, 13α-angeloyloxylupanine and 13α-tigloyloxylupanine present in the analysed seeds; 13α-hydroxylupanine, the second most abundant alkaloid in *L. albus*, was present only in few samples, while in others it was detected only in trace amounts (Fig. 5).

The most commonly occurring QA in all variants was lupanine making up ca. 64% of the total alkaloid content in the bitter (BIT) chemotype (Fig. 4.). A similar percentage (albeit at significantly lower ppm values as indicated on Fig. 3) was observed in the 'Frieda' parent (FRA) and the stacked allele recombinant (STA), 62.25% and 60.06%, respectively. In other variants (DIA, DSW, HET, PAU), the contribution of lupanine towards the total QA profile was reduced, ranging between 36.75 and 47.31%. The second most abundant QA in the bitter chemotype was 13α-hydroxylupanine (16.57% of the total alkaloid content), which trend was also observed in other accessions, except in the stacked allele recombinant (STA), where this QA was only present in trace amounts. The relationship between these two, most commonly occurring QAs was further investigated by calculating their ratios (Fig. 6). Figure 6 shows that in the haplotypes, where the reduced QA levels are conferred by the determinant

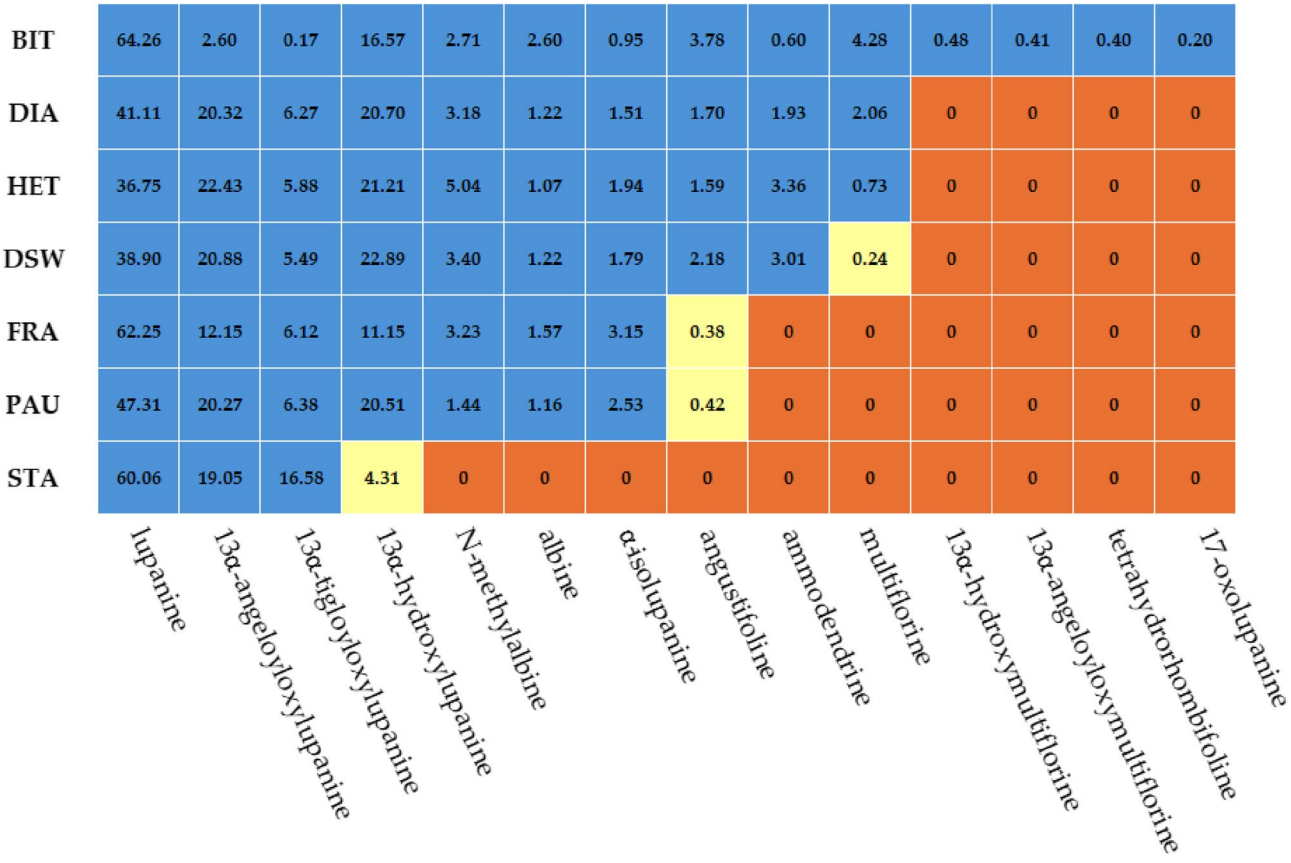


Fig. 5 Presence of QAs and their contribution in percentage in the total QA content in representatives of allelic combinations of *Lalb_Chr18_12359687* and *Lalb_Chr05_6643621*, ordered according to their abundance. A QA was assigned as present (blue cells) if its value was above 1 ppm (detection threshold of GC/FID), as present in trace amounts (yellow cell) when below 1 ppm (detection threshold of GC/FID), or as absent (red cells; GS/MS). BIT– bitter wild type progeny; FRA– 'Frieda' parent, homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DIA– 'Dieta' parent, wild type at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; PAU– homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DSW– wild type at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; HET– heterozygous at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; STA– homozygous sweet at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*

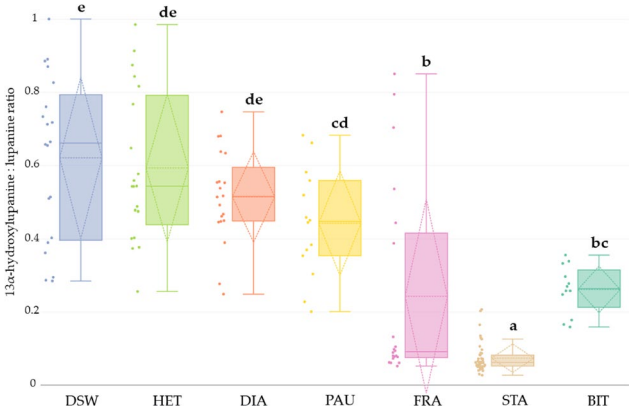


Fig. 6 Box plots of the ratio between the concentrations of 13α-hydroxylupanine and lupanine per variant. Values with the same letter are not significantly different at 0.05 confidence level. BIT– bitter wild type progeny; FRA– 'Frieda' parent, homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DIA– 'Dieta' parent, wild type at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; PAU– homozygous sweet at *pauper*, wild type at *Lalb_Chr05_6643621*; DSW– wild type at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; HET– heterozygous at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*; STA– homozygous sweet at *pauper*, homozygous sweet at *Lalb_Chr05_6643621*

on chromosome 5 (DIA, DSW, HET), the ratio between 13α-hydroxylupanine and lupanine was significantly higher compared to bitter (BIT) chemotype, while in stacked allele recombinant (STA), this ratio was significantly lower compared to all phenotypes.

Other, lupanine-derived QAs, 13α-angeloyloxylupanine was more abundant in the sweet chemotypes (12.15–22.43% of the total alkaloid content) compared to its presence (2.3% of the total alkaloid content) in the bitter one (Fig. 3.).

Discussion

We identified a new QTL associated to low QA content in white lupin, independent from *pauper* [14]. Furthermore, we found that the low QA associated haplotype in 'Dieta' was present (though inconsistently) in an old white lupin cultivar 'Start' which could indicate that the unidentified source of sweetness in 'Dieta' may be *exiguus* [30].

The newly identified QTL on chromosome 5 covers a large region (ca. 5.8%) of the chromosome. This could be due to the small bulk sizes, as other BSA studies include bulks of at least 20–50 individuals [31] to hundreds [32]. The limitation in our panel was created by the small sizes of the sampled subpopulations representing the phenotypic extremes, nonetheless highlighting the power of BSA even with small bulk sizes.

The QTL region consists of 10 genes with $-\log_{10}$(p-value) > 4.5, out of which the gene *Lalb_Chr05g0222381* was selected for marker development, based on its functional annotation and further investigations. The white lupin genome browser available on Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute [33], describes this gene to encode a "(1 of 3) K16276 - zinc finger protein-like protein (K16276, BTS)" with further information available about domains within. Amongst others, the functional annotation suggests a RING and CTCHY-type zinc finger domain within this gene, known transcriptional repressors of alkaloids in *Catharanthus roseus* [34]. On the white lupin genome browser [15], it encodes for a "Putative aminoacyltransferase, E1 ubiquitin-activating enzyme", which function is homologous to the enzyme LaHMT/HLT, an acetyltransferase responsible for forming acetylated products of 13α-hydroxylupanine and 13α-hydroxymultiflorine [35–37]. Although the gene *Lalb_Chr05g0222381* is associated to the low QA chemotype in the F_2 and F_3 populations, its involvement in the QA biosynthetic pathway cannot be concluded by our results. In fact, in the validation panel, the varieties 'Magnus', 'Orus', 'Luxe', and 'Clovis' are genotypically stacked allele recombinants (Supplementary Table S5) but their alkaloid content and profile does not match the stacked allele recombinants` (FiBL, unpublished data), suggesting that *Lalb_Chr05g0222381* might not be the causal gene for the low QA chemotype.

Our use of the proposed PACE marker was applied on F_2 and F_3 segregating populations. Since the markers are linked to but likely not causal, further studies are required to estimate the genetic distance to the sweetness QTL on chromosome 5 and hence their use and limitations for marker assisted selection. There is very rapid LD decay (< 4 kb) in the white lupin genome [15], reported both for germplasm panels carrying only sweet-seed inbred lines or genetically diverse materials including landraces and primitive populations. Nevertheless, whether this impacts the use of the markers as diagnostic tool, depends of the genetic distance between the markers and the causal locus for the sweetness QTL on chromosome 5. Furthermore, such fast LD decay may be beneficial in attempts on further pyramiding recently reported QTLs for anthracnose resistance (gene *Lalb_Chr05g0216161*) and protein content (gene *Lalb_Chr05g0224011*), located 3,685,681 bp upstream and 1,112,291 bp downstream the *Lalb_Chr05_6643621* locus, respectively [38, 39]. Based on the LD values reported for white lupin, physical distance of 1–3 Mbp between particular QTLs should be more than enough to break the linkage in breeding programs.

The *pauper*-associated gene, *Lalb_Chr18g0051511* encodes an acetyltransferase, which is responsible to block the QA biosynthetic pathway at an earlier stage, a step immediately downstream of the production of Δ1-piperideine by the enzyme CAO [14]. This change in biosynthesis results in a distinct chemotype in accessions that are homozygous for *pauper* sweetness lacking ammodendrine (the product of

the enzyme acetyltransferase) [14]. Our study confirmed the lack of ammodendrine in the *pauper* sweet parent (FRA) and in the *pauper* sweet progeny (PAU). Comparatively, individuals that only carry the allele associated to the parent 'Dieta' and sweetness on chromosome 5 (DIA, DSW, HET) exhibited a higher QA content in these accessions due to a distinct chemotype, characterized by the additional production of ammodendrine, angustifoline and multiflorine, along with a greater accumulation of 13-OH derivatives of lupanine (13α-hydroxylupanine, 13α-angeloyloxylupanine and 13α-tigloyloxylupanine). In stacked allele recombinants (STA), only lupanine and its esterified derivates, 13α-angeloyloxylupanine and 13α-tigloyloxylupanine, were present.

Kroc et al. [4] describe that the concentrations and their ratios of white lupin QAs have specific patterns. Lupanine was the most abundant alkaloid in their panel, with a lower share in sweet accessions, which was consistent with the data presented in this study. In fact, lupanine was in all recombinants the most abundant QA and its share in the total alkaloid profile remained ca. 60% in the bitter (BIT) chemotype, in the 'Frieda' parent (FRA) and in the stacked allele recombinant (STA). While the other sweet haplotypes, such as the *pauper* associated progeny (PAU) and individuals carrying only the allele associated to the parent 'Dieta' (DIA, DSW, HET), showed a reduction of lupanine content percentage ranged from 47.31% down to 36.75%. Furthermore, Kroc et al. [4] describe the presence of lupanine being complemented only by 13α-hydroxylupanine, with a lower share in low QA accessions, i.e. higher ratio of 13α-hydroxylupanine: lupanine. This observation was confirmed in our low QA individuals carrying only the allele associated to the parent 'Dieta' (sweetness on chromosome 5; DIA, DSW, HET) which showed a significant increase in the ratio of 13α-hydroxylupanine: lupanine compared to the bitter chemotype. In the stacked allele recombinant (STA) the ratio between 13α-hydroxylupanine: lupanine was very low, due to 13α-hydroxylupanine being present only in trace amounts. However, lupanine was complemented by its esterified QAs, 13α-angeloyloxylupanine and 13α-tigloyloxylupanine indicating that 13α-hydroxylupanine has been entirely converted into its esterified QAs. These findings suggest that the determinant on chromosome 5 is involved in the biosynthetic pathway branch that begins with lupanine leading to its 13-OH derivates, in particular in the key step that catalyzes the transformation of lupanine to 13α-hydroxylupanine. This hypothesis was consistent with the putative acyltransferase function assigned to this gene in the white lupin genome browser, though it might not be the causal gene [14, 36–38].

It should be noted that recombinants carrying stacked sweetness alleles of *pauper* and *Lalb_Chr05_6643621* revealed alkaloid levels (22.8 ± 10.4 ppm) much lower than previously reported minimal values for the European white lupin gene bank [4], including domesticated materials (95414, Bialorus-1, 159 ppm; 95413, Kiev Mutant, 160 ppm and 95234, P.21525, 168 ppm). However, low QA contents in lupins have been shown to correlate to increased susceptibility to aphid infestation. In blue lupin, the highest multiplication rate of aphids was associated to the accession with the lowest QA content of 64.1 ppm [40]. Moreover, some aphids (including at least one species, *Macrosiphum albifrons*, feeding on white lupin) are tolerant for alkaloids and even use them for their own defence [41]. Although no aphid infestation occurred in the F_2 and F_3 experimental populations, the response to aphids of the stacked allele recombinants as well as general agronomic performance should be addressed. Regardless, achieved progress in white lupin seed alkaloid content reduction is similar to that reported for narrow-leafed lupin, where minimal values reached 8–55 ppm in cultivars and 4-146 ppm in breeding lines [42–44]. Such a low alkaloid content as presented in white lupin *pauper* and *Lalb_Chr05_6643621* stacked recombinants should provide a safe margin for maintaining the recommended threshold of 200 ppm according to the variability of alkaloid content between years reported in sweet narrow-leafed lupin breeding lines [42].

Conclusion

The BSA in the F_2 population between the *pauper*-sweet cultivar 'Frieda' and the sweet by unknown source cultivar 'Dieta' allowed mapping of a new QTL for QA content in white lupin located on chromosome 5. Within this QTL, candidate genes were investigated to study genotype-phenotype associations. The QTL could be exploited in the future through fine mapping, identifying other candidate genes for potential validation, and further validated in a larger and more diverse panel of accessions. Regardless, the PACE marker tagging the SNP *Lalb_Chr05_6643621* associated to the gene *Lalb_Chr05g0222381* tagging the Dieta-type sweetness, in combination with the marker tagging *pauper* are powerful tools for breeders by taking advantage of two different sources of sweetness thanks to the ability to select for the trait of low QA content at early generations discriminating for bitter and heterozygous alleles (F_2-F_3). This can support the development of white lupin cultivars with low and stable QA content regardless of the growing environment.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06951-7.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
The authors would like to express their gratitude towards René Stefani for providing space for the field experiments, the FiBL Plant Breeding Team`s temporary employees across the years for the support during the fieldwork and David McNaughton (Soya UK Ltd; Southampton, UK) for providing seeds of the cultivar 'Dieta'. Furthermore, Magdalena Kroc (PAN-IGR; Poznań, Poland) and Chris Maliepaard (WUR; Wageningen, Netherlands) for their intellectual support along this study. Lastly, Prof. Dr. Michael Wink (University of Heidelberg, Germany) for kindly providing the analytical standard lupanine for the GC analysis.

Authors' contributions
A.P.: Writing– original draft, review, editing. Conceptualization. Data collection. Analysis. Visualization. Interpretation. Data curation.; M.Ka.– Writing– original draft, review, editing. Conceptualization. Data collection. Analysis. Visualization. Interpretation.; M.S.: Writing– review, editing. Analysis. Visualization. Interpretation. Data curation.; C.A.: Writing– review, editing. Data collection. Interpretation. Field trial design. Development of segregating populations. Acquisition.; M.M.M.: Writing– review, editing. Conceptualization. Acquisition.; E. B.: Writing– review, editing. Conceptualization. Data collection. Visualization. Interpretation.; A.T.: Writing– review, editing. Conceptualization. Data collection. Visualization. Interpretation.; M.Ks.: Writing– review, editing. M. L.: Writing– original draft, review, editing. Conceptualization. Data collection. Analysis. Visualization. Interpretation. Data curation. Supervision. Acquisition.All authors have read and approve the submitted version (and any substantially modified version that involves the author's contribution to the study) and have agreed to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Funding
This research has received funding from the projects DIVINFOOD (funded by the EU, grant no.101000383), LiveSeeding (funded by the EU, SERI and UKRI, grant no.101059872) and LUPINNO SUISSE (funded by the Swiss Federal Office for Agriculture - FOAG, contract: 2020/51).

Data availability
Sequencing data generated and/or analysed during the current study are available in the public ENA repository, with accession number PRJEB86989. All other data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL), Ackerstrasse 113, Frick 5070, Switzerland
[2]Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, Poznań 60-479, Poland
[3]Getreidezüchtung Peter Kunz (gzpk), Seestrasse 6, Feldbach 8714, Switzerland
[4]CREA-Research Centre for Animal Production and Aquaculture, Viale Piacenza 29, Lodi 26900, Italy

Received: 25 March 2025 / Accepted: 24 June 2025
Published online: 22 July 2025

References

1. Huyghe C. White lupin (Lupinus albus L). Field Crops Res. 1997;53(1–3):147–60.
2. Prusinski J. White lupin (Lupinus albus L) - Nutritional and health values in human nutrition - A review. Czech J Food Sci. 2017;35(2):95–105.
3. Frick KM, Lorensen MDBB, Esteban E, Pasha A, Schulz A, Provart NJ, et al. The aerial epidermis is a major site of quinolizidine alkaloid biosynthesis in narrow-leafed lupin [Internet]. Plant Biology; 2023 Mar [cited 2024 Jan 9]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.03.20.532575
4. Kroc M, Rybiński W, Wilczura P, Kamel K, Kaczmarek Z, Barzyk P, et al. Quantitative and qualitative analysis of alkaloids composition in the seeds of a white lupin (Lupinus albus L) collection. Genet Resour Crop Evol. 2017;64(8):1853–60.
5. Estivi L, Buratti S, Fusi D, Benedetti S, Rodríguez G, Brandolini A, et al. Alkaloid content and taste profile assessed by electronic tongue of Lupinus albus seeds debittered by different methods. J Food Compos Anal. 2022;114:104810.
6. Schreiber U, Schulte M, Khlopushina A, Esselen M. Hazard Identification of Food-Relevant Lupine Alkaloids Focusing on In Vitro Genotoxicity and Mutagenicity. ACS Food Sci Technol. 2025;5(2):726–33.
7. EFSA Panel on Contaminants in the Food Chain (CONTAM), Schrenk D, Bodin L, Chipman JK, del Mazo J, Grasl-Kraupp B, et al. Scientific opinion on the risks for animal and human health related to the presence of quinolizidine alkaloids in feed and food, in particular in lupins and lupin-derived products. EFSA J [Internet]. 2019 Nov [cited 2025 Feb 12];17(11). Available from: https://data.europa.eu/doi/10.2903/j.efsa.2019.5860
8. Brändle I, Arncken C, Kretzschmar U, Nicod L, Lazzaro M. Alkaloid analysis in lupins [Internet]. CH-Frick: Research Institute of Organic Agriculture FiBL; 2024 [cited 2025 Jul 5]. Available from: https://orgprints.org/id/eprint/52608/.
9. Bundesinstitut Für Risikobewertung. Risk assessment of the occurrence of alkaloids in lupin seeds. 2017 Mar 27 [cited 2025 Feb 28]; Available from: http://www.bfr.bund.de/cm/349/risk-assessment-of-the-occurrence-of-alkaloids-in-lupin-seeds.pdf
10. EUROPEAN COMMISSION. Standing Committee on Plants, Animals, Food and Feed Section Novel Food and Toxicological Safety of the Food Chain [Internet]. European Commission Health and Food Safety Directorate General; 2024 [cited 2025 Mar 18]. Report No.: sante.g.3(2024)10312859. Available from: https://food.ec.europa.eu/document/download/acdba393-267f-4c68-9cea-84448a7288ee_en?filename=reg-com_toxic_20240920_sum.pdf&prefLang=ga
11. Harrison JEM, Williams W. Genetical control of alkaloids in Lupinus albus. Euphytica. 1982;31(2):357–64.
12. Sengbusch RV. The breeding of sweet lupins. Herbage Rev. 1938;6(2):1–9.
13. Hondelmann W. The lupin — ancient and modern crop plant. Theor Appl Genet. 1984;68–68(1–2):1–9.
14. Mancinotti D, Czepiel K, Taylor JL, Golshadi Galehshahi H, Møller LA, Jensen MK, et al. The causal mutation leading to sweetness in modern white lupin cultivars. Sci Adv. 2023;9(31):eadg8866.
15. Hufnagel B, Soriano A, Taylor J, Divol F, Kroc M, Sanders H, et al. Pangenome of white lupin provides insights into the diversity of the species. Plant Biotechnol J. 2021;19(12):2532–43.
16. Plarre W. Lupinen. In: Fischbeck G, Plarre W, Schuster W, editors. Lehrbuch der Züchtung landwirtschaftlicher Kulturpflanzen. 2nd ed. Stuttgart: Paul Parey; 1985. p. 185–96.
17. Li X, Yang H, Buirchell B, Yan G. Development of a DNA marker tightly linked to low-alkaloid gene iucundus in narrow-leafed lupin (Lupinus angustifolius L) for marker-assisted selection. Crop Pasture Sci. 2011;62(3):218.
18. Wink M, Meißner C, Witte L. Patterns of quinolizidine alkaloids in 56 species of the genus Lupinus. Phytochemistry. 1995;38(1):139–53.
19. Resta D, Boschin G, D'Agostina A, Arnoldi A. Evaluation of total quinolizidine alkaloids content in lupin flours, lupin-based ingredients, and foods. Mol Nutr Food Res. 2008;52(4):490–5.
20. Boschin G, Annicchiarico P, Resta D, D'Agostina A, Arnoldi A. Quinolizidine Alkaloids in Seeds of Lupin Genotypes of Different Origins. J Agric Food Chem. 2008;56(10):3657–63.
21. Linstrom, Peter, Mirokhin Y, Tchekhovskoi D, Yang, Xiaoyu. NIST/EPA/NIH Mass Spectral Library (NIST 14) and NIST Mass Spectral Search Program (Version 2.2) [Internet]. Gaithersburg, MD: U.S. Department of Commerce National Institute of Standards and Technology; 2014. Available from: https://www.nist.gov/system/files/documents/srd/NIST1aVer22Man.pdf

22. Schneider M, Shrestha A, Ballvora A, Léon J. High-throughput estimation of allele frequencies using combined pooled-population sequencing and haplotype-based data processing. Plant Methods. 2022;18(1):34.
23. Andrews S. FastQC [Internet]. 2010. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
24. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. arXiv; 2013 [cited 2025 Jan 10]. Available from: http://arxiv.org/abs/1303.3997
25. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015;31(12):2032–4.
26. Gautier M. Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. Genetics. 2015;201(4):1555–79.
27. R: The R Project for Statistical Computing [Internet]. [cited 2025 Feb 28]. Available from: https://www.r-project.org/
28. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means [Internet]. 2017 [cited 2025 Mar 5]. p. 1.10.7. Available from: https://CRAN.R-project.org/package=emmeans
29. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015;67:1–48.
30. Rychel S, Książkiewicz M. Development of gene-based molecular markers tagging low alkaloid pauper locus in white lupin (Lupinus albus L). J Appl Genet. 2019;60(3–4):269–81.
31. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, et al. QTL -seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 2013;74(1):174–83.
32. Xin F, Zhu T, Wei S, Han Y, Zhao Y, Zhang D, et al. QTL Mapping of Kernel Traits and Validation of a Major QTL for Kernel Length-Width Ratio Using SNP and Bulked Segregant Analysis in Wheat. Sci Rep. 2020;10(1):25.
33. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40(Database issue):D1178-86.
34. Pauw B, Hilliou FAO, Martin VS, Chatel G, De Wolf CJF, Champion A, et al. Zinc Finger Proteins Act as Transcriptional Repressors of Alkaloid Biosynthesis Genes in Catharanthus roseus. J Biol Chem. 2004;279(51):52940–8.
35. Suzuki H, Murakoshi I, Saito K. A novel O-tigloyltransferase for alkaloid biosynthesis in plants Purification, characterization, and distribution in Lupinus plants. J Biol Chem. 1994;269(22):15853–60.
36. Okada T, Hirai MY, Suzuki H, Yamazaki M, Saito K. Molecular Characterization of a Novel Quinolizidine Alkaloid O-Tigloyltransferase: cDNA Cloning, Catalytic Activity of Recombinant Protein and Expression Analysis in Lupinus Plants. Plant Cell Physiol. 2005;46(1):233–44.
37. Saito K, Suzuki H, Takamatsu S, Murakoshi I. Acyltransferases for lupin alkaloids in Lupinus hirsutus. Phytochemistry. 1992;32(1):87–91.
38. Schwertfirm G, Schneider M, Haase F, Riedel C, Lazzaro M, Ruge-Wehling B, et al. Genome-wide association study revealed significant SNPs for anthracnose resistance, seed alkaloids and protein content in white lupin. Theor Appl Genet. 2024;137(7):155.
39. Alkemade JA, Nazzicari N, Messmer MM, Annicchiarico P, Ferrari B, Voegele RT, et al. Genome-wide association study reveals white lupin candidate gene involved in anthracnose resistance. Theor Appl Genet. 2022;135(3):1011–24.
40. Philippi J, Schliephake E, Jürgens HU, Jansen G, Ordon F. Correlation of the alkaloid content and composition of narrow-leafed lupins (Lupinus angustifolius L) to aphid susceptibility. J Pest Sci. 2016;89(2):359–73.
41. Philippi J, Schliephake E, Jürgens HU, Jansen G, Ordon F. Feeding behavior of aphids on narrow-leafed lupin (Lupinus angustifolius) genotypes varying in the content of quinolizidine alkaloids. Entomol Exp Appl. 2015;156(1):37–51.
42. Kamel KA, Święcicki W, Kaczmarek Z, Barzyk P. Quantitative and qualitative content of alkaloids in seeds of a narrow-leafed lupin (Lupinus angustifolius L) collection. Genet Resour Crop Evol. 2016;63(4):711–9.
43. Khedr T, Gao LL, Kamphuis LG, Bose U, Juhász A, Colgrave ML. Evaluation of alkaloid levels in commercial and wild genotypes of narrow-leafed lupin. J Food Compos Anal. 2024;135:106600.
44. Kroc M, Koczyk G, Kamel KA, Czepiel K, Fedorowicz-Strońska O, Krajewski P, et al. Transcriptome-derived investigation of biosynthesis of quinolizidine alkaloids in narrow-leafed lupin (Lupinus angustifolius L) highlights candidate genes linked to iucundus locus. Sci Rep. 2019;9(1):2231.

Publisher's note

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

Supplementary information

Supplementary Tables in Publication nr. 1 (Patyi et al., 2025) include:

Supplementary Table S1: List of F2 progeny derived from the 'Dieta' x 'Frieda' crossing, grown in 2023, indicating if it was carried over to F3, the individual phenotypic score from the Dragendorff test and the allelic state at the relevant loci.

Supplementary Table S2: List of F3 progeny derived from the 'Dieta' x 'Frieda' crossing, grown in 2023, indicating the F2 parental line, the individual phenotypic score from the Dragendorff test and the allelic state at the relevant loci.

Supplementary Table S3: List of accessions used in the validation panel and the re-sequenced accessions available on whitelupin.fr indicating their status, sweetness status and allelic states at relevant loci.

Supplementary Table S4: Alkaloid content in the indicated subset of individuals from F3, including biological and technical replicates, allelic states at pauper and Lalb_Chr05_6643621, individual concentrations of the 14 analysed QAs, total alkaloid content (in ppm) and the ratio between 13a-hydroxylupanine and lupanine; BIT – bitter wild type progeny; FRA – 'Frieda'parent, homozygous sweet at pauper; DIA – 'Dieta' parent, homozygous sweet at Lalb_Chr05_6643621; PAU – homozygous sweet at pauper, wild type at Lalb_Chr05_6643621; DSW – wild type at pauper, homozygous sweet at Lalb_Chr05_6643621; HET – heterozygous at pauper, homozygous sweet at Lalb_Chr05_6643621; STA – homozygous sweet at pauper, homozygous sweet at Lalb_Chr05_6643621

Supplementary Table S5: Sequencing data infromation from the raw reads per bulk (read length of 150 bp), after quality filtering, mapping and removal of redundancies

Supplementary Table S6: Genes and associated -log10(p-value) based on differences in allele frequencies of haplotypes between two bulks of the F2 population representing high and low QA levels

Supplementary Table S7: Genes with -log10(p-value) > 4.5 associated to the non-pauper sweet haplotype in the QTL region spanning from 5799140 bp to 7304977 bp on chromosome 5.

All supplementary tables are available in the original published scientific article (https://link.springer.com/article/10.1186/s12870-025-06951-7?utm_source=researchgate.net&utm_medium=article#Sec23) and available for download at https://static-content.springer.com/esm/art%3A10.1186%2Fs12870-025-06951-7/MediaObjects/12870_2025_6951_MOESM1_ESM.xlsx

Declaration of the PhD student`s contribution in publications and the authors` declarations of their contribution

Publication nr. 1

Patyi, A., Kamp, M., Arncken, C., Biazzi, E., Książkiewicz, M., Messmer, M. M., Schneider, M., Tava, A., & Lazzaro, M. (2025). Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (Lupinus albus, L.) and development of ultra-low alkaloid recombinants by stacking with the pauper allele. BMC Plant Biology, 25(1), 945. https://doi.org/10.1186/s12870-025-06951-7

András Patyi, M.Sc.

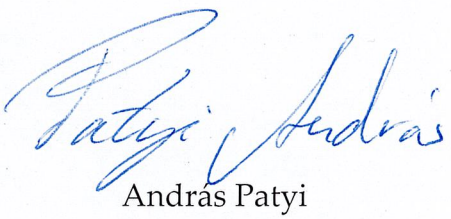
Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
5070 Frick
Switzerland

Institute of Plant Genetics, Polish Academy of Sciences
Strzeszyńska 34
60-479 Poznań
Poland

Declaration

I, András Patyi, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system, as well as performed the duties of corresponding author:

"A.P.: Writing – original draft, review, editing. Conceptualization. Data collection. Analysis. Visualization. Interpretation. Data curation."



András Patyi

Frick, 2026.04.22.

Miriam Kamp, M.Sc.

Getreidezüchtung Peter Kunz (gzpk)
Seestrasse 6
8714 Feldbach
Switzerland

Declaration

I, Miriam Kamp, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (Lupinus albus, L.) and development of ultra-low alkaloid recombinants by stacking with the pauper allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"M.Ka.: Writing– original draft, review, editing. Conceptualization. Data collection. Analysis. Visualization. Interpretation."

Miriam Kamp

Miriam Kamp

Feldbach, 2026.04.24.

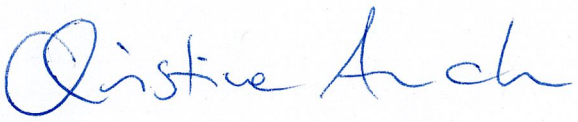
Christine Arncken, Dipl.-Ing. agr. ETH

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
Frick 5070
Switzerland

Declaration

I, Christine Arncken, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"C.A.: Writing– review, editing. Data collection. Interpretation. Field trial design. Development of segregating populations. Acquisition."



Christine Arncken

Frick, 2026.04.23.

Elisa Biazzi, Dr.

CREA-Research Centre for Animal Production and Aquaculture
Viale Piacenza 29
26900 Lodi
Italy

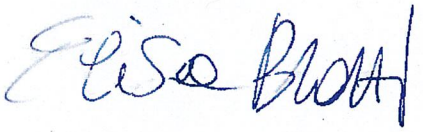
Declaration

I, Elisa Biazzi, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"E. B.: Writing– review, editing. Conceptualization. Data collection. Visualization. Interpretation."

Elisa Biazzi

Lodi, 2026.04.23.



dr hab. Michał Książkiewicz, prof. IGR PAN

Institute of Plant Genetics, Polish Academy of Sciences
Strzeszyńska 34
60-479 Poznań
Poland

Declaration

I, Michał Książkiewicz, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"M.Ks.: Writing– review, editing."

Michał Książkiewicz

Poznań, 2026.04.29.

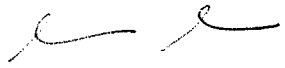
Monika M. Messmer, Dr.

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
5070 Frick
Switzerland

Declaration

I, Monika Messmer, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"M.M.M.: Writing– review, editing. Conceptualization. Acquisition."



Monika M. Messmer

Frick, 2026.04.22.

Michael Schneider, Dr.

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
5070 Frick
Switzerland

Declaration

I, Michael Schneider, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"M.S.: Writing– review, editing. Analysis. Visualization. Interpretation. Data curation."



Michael Schneider

Frick, 2026.04.23.

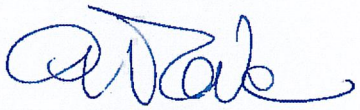
Aldo Tava, Dr.

CREA-Research Centre for Animal Production and Aquaculture
Viale Piacenza 29
26900 Lodi
Italy

Declaration

I, Aldo Tava, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"A.T.: Writing– review, editing. Conceptualization. Data collection. Visualization. Interpretation."



Aldo Tava

Lodi, 2026.04.23.

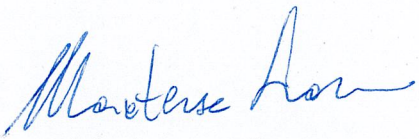
Mariateresa Lazzaro, Dr.

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
Frick 5070
Switzerland

Declaration

I, Mariateresa Lazzaro, declare that in the publication titled "Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the *pauper* allele", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"M. L.: Writing– original draft, review, editing. Conceptualization. Data collection. Analysis. Visualization. Interpretation. Data curation. Supervision. Acquisition."



Mariateresa Lazzaro

Frick, 2026.04.22.

Publication nr. 2

Patyi, A., Schneider, M., Arncken, C., Książkiewicz, M., Messmer, M. M., Schwertfirm, G., Lazzaro, M. (2026). From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments. *Mol Breeding* 46, 37. https://doi.org/10.1007/s11032-026-01661-w

Journal metrics:

- Impact Factor (2024): 3.0
- 5-year journal impact factor (2024): 3.0
- MNiSW: 70

Molecular Breeding (2026) 46:41
https://doi.org/10.1007/s11032-026-01661-w

Molecular Breeding

RESEARCH

From research to application: evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments

András Patyi[1,2] · Michael Schneider[1] · Christine Arncken[1] · Michał Książkiewicz[2] · Monika M. Messmer[1] · Grit Schwertfirm[3] · Mariateresa Lazzaro[1]

Received: 11 December 2025 / Accepted: 6 April 2026
© The Author(s), under exclusive licence to Springer Nature B.V. 2026

Abstract

Resistance to anthracnose disease, caused by *Colletotrichum lupini*, in white lupin is polygenic and rare in global germplasm, making breeding for resistance challenging. Several studies have identified genomic regions associated with resistance, but reported QTLs rarely overlap. Overall, anthracnose-associated loci have not yet been appropriately exploited in breeding due to the lack of validation, especially in unrelated material. We assessed three literature-based SNP-markers associated to a previously described QTL, as well as six originating from two published GWAS studies. In addition, by re-analysing two published datasets, we newly (i) identified five SNP-markers through a genome-wide association study (GWAS) and (ii) assembled a genomic prediction (GP) model based on a set of 42 SNP-markers. All SNP-markers were transformed to PCR-based assays to facilitate cost-efficient application in breeding and tested in a diversity panel and two breeding panels, phenotyped in field and/or controlled conditions. We validated the association to anthracnose resistance for the previously reported QTL on chromosome *Lalb_Chr10*, tagging a source of resistance in the resistant cultivar 'Frieda'. Two markers showed significant associations ($p<0.05$) with anthracnose resistance in at least two panels, and an additional seven markers were significant in one. The GP model had a prediction accuracy of 0.84 (± 0.13) in the training panel, but is not applicable as such in the unrelated validation panels. However, re-training the GP model with the respective validation panel data showed clear prediction improvement buffering the limitation of environmental context and narrower genepools applied in a breeding programme.

Extended author information available on the last page of the article

Published online: 22 April 2026

Keywords White lupin · Anthracnose · Marker assisted selection (MAS) · Genome-wide association study (GWAS) · Genomic prediction (GP) · Molecular breeding

Introduction

White lupin (*Lupinus albus* L.; diploid, 2n = 50) is a Mediterranean grain legume with high potential as a protein source for feed and food (Hufnagel et al. 2020). However, its cultivation is severely constrained by anthracnose, caused by the ascomycete *Colletotrichum lupini*. All lupin species are susceptible to this disease, which is particularly destructive in temperate regions of Europe. Anthracnose infects all aerial plant parts, impairing vegetative growth and pod development, and can result in complete yield loss in unfavourable years (Alkemade et al. 2025).

The major outbreak of this fungal pathogen, with suggested centre of origin in Latin America (Alkemade et al. 2021a, b, 2023), dates in the 1970 s and 80 s in Europe and South America. It quickly became a worldwide pandemic destroying lupin fields, especially since the commonly used commercial varieties at the time showed to be susceptible to this disease (Talhinhas et al. 2016). Thus, the new objective of plant breeders was to develop resistant varieties against *C. lupini*. For white lupin, a large screening effort was conducted in Australia between 1997 and 1998, assessing anthracnose resistance in ca. 300 genetic resources of white lupin from across the globe (Adhikari et al. 2009, 2013). The only identified source of resistance was found in an Ethiopian genepool of landraces collected in the highland region (1900–2600 m altitude) in the provinces of Gonder and Gojam. The most resistant accessions were collected from the Eastern Gojam region (Debre Markos district), including accessions P27174, P27175, and P27178 (Adhikari et al. 2009). More recently, a large number of genetic resources (gene bank accessions, landraces, cultivars and breeding lines) were evaluated in the field by FiBL (approximately 100 accessions per year from 2015 to 2022, 890 in total; unpublished data). Accessions were screened once late in the growing season (BBCH82-87) for their disease scores ranging from 1 to 9 (1 – healthy, 9 – dead, as described by Alkemade et al. 2021a, b), grown with an infection row on each side. No accessions showed complete resistance against the pathogen, and field performance further confirmed the rarity of increased resistance in the global white lupin germplasm.

Motivated by the severe impact of anthracnose on white lupin cultivation, numerous molecular studies have investigated the genetic basis of resistance. Early work focused on a bi-parental F_8 recombinant inbred line (RIL) population derived from the cross `Kiev mutant` (susceptible, CPVO, 2026)× P27174 (resistant, Adhikari et al. 2009). Using this population, Phan et al. (2007) identified two major quantitative trait loci (QTLs) on linkage groups LG04 and LG17, corresponding to current chromosomes *Lalb_Chr04* and *Lalb_Chr02*, which together explained more than half of the phenotypic variance (31% and 26% respectively). These results suggested that anthracnose resistance is a quantitative trait under polygenic control. Subsequent analyses of the same RIL population refined this genetic framework and proved the trait`s polygenic nature. Yang et al. (2010) mapped three resistance-associated mark-

ers (WANR1 and WANR2 on Lalb_Chr02, and WANR3 on *Lalb_Chr04*) together explaining 59% of phenotypic variation. The development of a high-density consensus linkage map (Książkiewicz et al. 2017) further confirmed QTLs on *Lalb_Chr02* and *Lalb_Chr04* and identified an additional locus on *Lalb_Chr10*. Integrating these findings, Rychel-Bielska et al. (2020) evaluated the available markers for breeding applications and developed three CAPS markers suitable for marker-assisted selection (MAS) targeting the QTLs on *Lalb_Chr02* and *Lalb_Chr04*, while reporting limited success of the WANR markers in breeding programs. Despite these advances, the transferability of resistance-associated markers remained a challenge. A genomic prediction model developed by Alkemade et al. (2021a, b), based on the markers reported by Rychel-Bielska et al. (2020), showed no correlation between predicted and observed resistance values. Notably, the resistant parental line of the RIL population (P27174) exhibited an opposite anthracnose resistance pattern in the 2021 study (Alkemade et al. 2021a, b) compared to that displayed in their original phenotyping in Australia (Adhikari et al. 2009), underscoring the strong influence of environment and experimental conditions.

The availability of a reference genome for the cultivar Amiga (CPVO, 2026), together with resequencing data from 39 accessions (Hufnagel et al. 2020, 2021), enabled a shift from bi-parental mapping to genome-wide association studies (GWAS) in more diverse germplasm. Alkemade et al. (2022) analysed a panel of 181 accessions using 4,611 SNPs and phenotyping with stem-wound inoculation method in controlled conditions. This GWAS identified three closely linked SNPs on *Lalb_Chr05* associated with anthracnose resistance and reported additional resistant accessions, including material from Ethiopia, two Chilean breeding lines, and the newly released cultivar `Frieda` (CPVO, 2026).

However, the robustness of these associations was challenged by later work. Schwertfirm et al. (2024) conducted GWAS and genomic prediction using 24,576 SNPs in a panel of 255 accessions phenotyped across multiple years, locations, and developmental stages under field condition. This study could not validate the *Lalb_Chr05* associations, nor the resistance of the previously reported Chilean accession, a result corroborated by independent field trials in Switzerland (FiBL, unpublished data). Instead, Schwertfirm et al. (2024) identified 17 significant SNPs distributed across 14 chromosomes. Only three of these overlapped with loci previously reported in the `Kiev mutant` × P27174 population, corresponding to known QTLs (*Lalb_Chr04*, *Lalb_Chr10*).

Collectively, these studies highlight the major challenges facing anthracnose resistance breeding in white lupin. Resistance is scarce in global germplasm, strongly influenced by phenotyping methodology and environment, and controlled by multiple loci with limited consistency across studies. The poor overlap between reported QTLs and GWAS signals has hindered the effective and consistent deployment of marker-assisted selection. Consequently, a key remaining challenge is the validation of cost-efficient molecular assays that reliably tag the respective resistance loci across independent gene pools, with particular emphasis on breeding-relevant material. Addressing this gap is essential for translating molecular discoveries into practical resistance breeding strategies.

The present study aims to address this research to application gap by (i) exploiting already published datasets for further loci identification with GWAS, and for developing a GP model based on a minimal number of markers as a blue-print for low cost application; (ii) merging all suitable literature and newly identified markers to one marker assay-type, PCR Allelic Competitive Extension (PACE), tagging reference and alternative alleles with distinct fluorescent signals; (iii) testing these SNP-markers in new genetic pools (namely one diversity panel and two breeding panels) and against different phenotyping approaches (field and controlled conditions); (iv) applying a GP model retraining approach to improve performance in the selection of breeding materials.

Overall, this study develops and evaluates SNP marker assays for application in anthracnose resistance selection, based on the hypothesis that switching from marker-trait discovery to application can be limited by (i) a different, and in some cases narrower genepool and (ii) the origin of the phenotypic approach in the discovery phase compared to the validation.

Materials and methods

Plant material

Public datasets

The LUW (LUpin White, $n = 255$) panel (downloaded from the supplementary material of Schwertfirm et al. 2024), consists of 167 gene bank accessions, 73 advanced breeding lines from the Landwirtschaftliche Lehranstalten Triesdorf (LLT, Germany) and 15 commercially available cultivars (Table 1).

The panel from Alkemade et al. 2022 (downloaded from the supplementary material of the study) consists of 200 commercially available cultivars, landraces and breeding lines (training panel) with enough phenotypic data available for 192 of them (Table 1).

New diversity and breeding panels

The genetic pools (Table 1) newly phenotyped in this study to validate the consistency of individual marker effects and the accuracy of genomic prediction are (i) a diversity panel, curated with the intention to represent diverse genetic backgrounds and diverse sources of resistance, phenotyped in controlled conditions; (ii) a multi-parental panel of F_4-F_6 breeding lines, phenotyped in controlled and field conditions; (iii) a biparental panel consisting of F_3 individuals with low quinolizidine alkaloid content. The latter two are breeding panels from the common FiBL and gzpk (*Getreidezüchtung Peter Kunz*, Feldbach, Switzerland) breeding programme (Table 1).

Table 1 Description of the panels used, including panel name, type, material composition, number of individuals and the phenotyping environment(s) with information on the phenological stage of the plants

Panel	Type	Use	Materials	*n*	Phenotyping
LUW	Public	New GWAS	landraces breeding lines cultivars	255	field adult (3 phenological stages) (Schwertfirm et al. 2024)
Training	Public	New GP	landraces breeding lines cultivars	192	climate chamber seedling (2–4 weeks-old) (Alkemade et al. 2022)
Diversity	New	Marker validation	wild type landraces breeding lines cultivars	48	climate chamber seedling (2–4 weeks-old) experiment in 2022
Multi-parental	New	Marker validation	breeding lines (F_4-F_6)	20 + 2[a]	climate chamber seedling (2–4 weeks-old) experiment in 2024
				94 + 2[a]	field adult (3 scorings) experiment in 2024
Biparental	New	Marker validation	breeding population ($F_{3)}$	50	field adult (1 late-season scoring) experiment in 2024

[a] The first number indicated the size of the validation panel, the second represents the cultivars `Frieda` and `Amiga`, which were added to the validation populations for reference

The **diversity panel** includes 48 accessions consisting of 20 breeding lines, 23 commercial cultivars, 4 landraces and the wild type *graecus* (Table 1, Supplementary Table S1). The **multi-parental panel** consists of 94 individuals, selected from 25 advanced breeding lines (3–5 individual plants of each, generation from F_4 to F_6), and two commercial cultivars used as controls for resistance (`Frieda`) and susceptibility (`Amiga`). All single plants were pheno- and genotyped individually, thus the entire panel represented 96 individuals. Their simplified pedigree, generation, disease score and allelic states are presented in Supplementary Table S2. All 94 individual plants from the 25 breeding lines were analysed in field experiments, whereas in the controlled environment, one representative on only 20 of these lines were analysed due to limited seed availability (by sowing 6 seeds from each plant) (Supplementary Table S3).

The **biparental panel** consisted of 50 individuals (Table 1) in F_3 generation, originating from an elite cross between cultivars `Frieda` x `Dieta` (Supplementary Table S4, CPVO, 2026) exhibiting exceptionally low quinolizidine alkaloid content through two independent loci associated low-alkaloid chemotype (Patyi et al. 2025). Segregating for resistance to anthracnose, this population served the purpose of the application of these molecular markers for selection of individuals with increased resistance.

New phenotypic scoring and analysis

Experiments in field conditions were conducted for the multi-parental and biparental panels in 2024 (Table 1, Supplementary Table S2, Supplementary Table S4), whereas in controlled conditions for diversity panel in 2022 (Table 1, Supplementary Table S1) and multi-parental panel in 2024 (Table 1, Supplementary Table S3).

In field experiments, each breeding line in the multi-parental panel was grown in unreplicated rows consisting of six individuals, surrounded by infection rows on both sides. The commercial cultivars `Frieda` and `Amiga` were repeated three times across the trial. The individuals of the biparental panel were grown without infection rows in the same field with high anthracnose infection pressure. Individuals of both breeding panels were grown under protective nets to prevent cross-pollination in Full-Reuenthal, Switzerland (47°35'59.5"N 8°12'10.4"E) in 2024. Phenotypic scores of the multi-parental (Supplementary Table S2) and the biparental panel (Supplementary Table S4) from the field correspond to scores of individual plants collected during the growing season. For the multi-parental panel, individuals were scored three times across the growing season (102, 120 and 128 days after sowing) to calculate standard area under disease progression curve (sAUDPC) scores (calculated using the package *agricolae* version 1.3–7.3 (Mendiburu 2023) in R version 4.4.2 (R Core Team, 2024), while the representatives of the biparental panel were scored only once in the field (BBCH82-87), to represent a less labour-intensive and more realistic phenotyping approach for early generation material.

In controlled environment experiments, artificial inoculation was performed as presented in Alkemade et al. (2021a, b) using the highly virulent *C. lupini* JA01 strain, isolated from Mellikon, Switzerland, in 2018. One individual plant representative for each breeding line in the multi-parental panel was selected, and 6 seeds of this plant were grown and phenotyped in the climate chamber. Each plant was assessed for disease symptoms at 0, 3, 7, 10 and 14 days post-inoculation to calculate sAUDPC scores (calculated using the package *agricolae* version 1.3–7.3 (Mendiburu 2023) in R version 4.4.2 (R Core Team, 2024). The adjusted sAUDPC scores of the offspring of the selected plants were correlated (Pearson correlation) to the sAUDPC values of the mother plants observed in the field.

The disease score for phenotypic values originating from controlled conditions was analysed using Linear Mixed Models (LMMs), with sAUDPC score as response variable, corresponding accession as fixed, and individual plant replication ($n=6$) nested within the climate chamber experiment as a random effect for the re-analysis of the training panel and the phenotypic values for all retrained models with phenotypic values originating from independent climate chamber experiments, and only biological replication ($n=6$) for the diversity- and multi-parental panels where phenotypic values originated from a single climate chamber experiment (Supplementary Table S1, Supplementary Table S3).

Estimated marginal means from the LMMs were calculated using *emmeans* version 1.11.1 (Lenth et al. 2025), which values were used for downstream analysis. Meanwhile, for the field observations, individual disease scores were used for the downstream analysis as individual plants were genotyped and phenotyped.

Broad sense heritability (H^2) in the validation panels as described by Schmidt et al. (2019) using the package inti version 0.6.9.1 (R Core Team, 2024). It was not possible to calculate H^2 values (Cullis) for the field datasets (multi-parental and biparental panels field observations) due to the unavailability of biological replication.

SNP markers

PACE assay design and cycling conditions

All loci from literature, as well as the newly identified ones had associated SNP loci, with known information of the location and available genomic context sequence on the white lupin genome browser (Hufnagel et al. 2020, 2021). DNA extraction was performed using Qiagen FTA cards (QIAGEN GmbH; Hilden, Germany). Primer triplets were designed by 3CR Bioscience (Essex, UK) based on the provided context sequence (±100 bp) for each polymorphic locus, with FAM and HEX fluorophores tagging the reference and alternative allele, respectively. For each PACE assay, the sequences of the two forward and the common reverse primers are presented in Supplementary Table S5 (markers that were used individually for the evaluation), and in Supplementary Table S6 (markers used for GP model evaluation). PCRs were performed using the newly developed primer triplets, with cycling conditions as described in Patyi et al. (2025) ran on the CFX Opus 384 Real-Time PCR System (Bio-Rad Laboratories, Inc.; Hercules CA, USA). Allelic calls were assigned automatically using CFX Maestro version 2.2 (Bio-Rad Laboratories, Inc.; Hercules CA, USA) which were used for further analysis.

Individually assessed SNP markers

The nine markers from literature include: (i) three from Rychel-Bielska et al. 2020, (ii) one from Alkemade et al. 2022 and (iii) five from Schwertfirm et al. (2024).

CAPS markers from Rychel-Bielska et al. (2020) tagging the QTLs (*Lalb_Chr02*, *Lalb_Chr04*) identified by Yang et al. (2010) were transformed as PACE marker assays. The context sequences of the three CAPS markers were downloaded on the respective chromosomes based on amplicon sequences (amplicon ID: TP222136, TP47110, TP338761; sequence retrieved from DDBJ, Altschul et al. 1997). In all three cases, the amplicon-associated sequence contained a SNP, which polymorphism was used for marker development.

Alkemade et al. 2022 identified three SNPs associated to anthracnose resistance. One out of the three, Lalb_Chr05_3706534 has only been associated to relative lesion size, which phenotypic measurement was not carried out in this study, therefore it was not considered for PACE assay development. The remaining two, Lalb_Chr05_2957601 and Lalb_Chr05_2957940 are closely linked (ca. 300 bp apart) being positioned on the coding sequence of the same gene on chromosome *Lalb_Chr05*. For this reason, in our study we used only SNP Lalb_Chr05_2957601.

Schwertfirm et al. (2024) report 17 SNPs associated to anthracnose resistance, from which the seven high confidence loci were considered for marker development. For Lalb_Chr19_2669672, PACE primer design was unfeasible due to the presence of other SNPs in the 50 bp flanking regions, while the SNP-marker Lalb_Chr04_16717968 failed during the PCR amplification. Thus, we used only five markers published by Schwertfirm et al. (2024) for evaluation in this study.

Furthermore, through the re-analysis of the datasets presented in Schwertfirm et al. (2024), we present two marker loci that overlap with the ones published in the original study, and present five additional markers, derived from the new GWAS analysis (Table 2, Supplementary Table S5). To complement the MMLM, FarmCPU and BLINK models presented in Schwertfirm et al. (2024), we performed the GWAS (raw data as provided in the original study) in R version 4.2.1. (R Core Team, 2022) using GAPIT3.3 (Wang and Zhang 2021) GLM function, including all markers, regardless of MAF and missing rate, to determine additional potential candidate loci. Including missing markers should enable the detection of "hidden" QTLs, as described by Gabur et al. (2018). All potential candidate SNPs have been assessed for relevance in the validation panels with PACE markers to account for the increased false positive detection rate of the GLM model.

After GWAS, the seven most significant SNPs with highest explained variance values ranging between 0.32 and 0.65 (Table 2) across all phenotypic data available (three phenological stages from field scorings) were selected for PACE assay development, irrespective of missing marker information. Two markers were overlapping with the ones presented in the original publication, while five additional ones were defined due to the additional models used and due to looser selection criteria (Supplementary Table S5).

Allelic states at the QTL/GWAS derived loci (Table 2) were analysed individually to determine significant association of the respective phenotypic values with analysis of variance (ANOVA) and post-hoc TukeyHS using the package stats version 3.6.2 (R Core Team, 2024) or with t-test to test the statistical difference ($p < 0.05$), depending on whether there are three or two allelic groups at the tested locus. Explained variance values were calculated independently for each marker locus as the proportion of the total sum of squares (SS_{total}) attributable to the marker (SS_{marker}), expressed as a percentage:

$$Explained\ Variance\ (\%) = \frac{SS_{marker}}{SS_{total}} \times 100$$

For the estimation of overall performance of the set of markers individually associated to the phenotype (Table 2) phenotypic variance explained (PVE) values were calculated and compared. PVE values for all validation panels were calculated (adjusted R^2 approach) using the package CJAMP version 0.1.1 (R Core Team, 2024), using a SNP only if it was polymorphic in the respective dataset.

Table 2 SNP-markers applied in this study for individual assessment. Locus of the target SNP on the respective chromosome, its status indicates if a specific SNP is newly described in this study or previously published, and the method with which it was identified. For de novo markers, p-values and explained variance is presented from the repeated GWAS analysis, specifying which model(s) describe the marker and which phenological stage`s phenotypic data (accessed from Schwertfirm et al. 2024) is it associated to, and how many accessions were genotyped (n)

Chromosome	Position	Dataset	Status[a]	Method	Phenotyping	*p*-value	Explained Variance (%)	*n*	Model
Lalb_Chr02	550,648	Rychel-Bielska et al. 2020	published	QTL mapping	field				
Lalb_Chr02	635,857	Rychel-Bielska et al. 2020	published		field				
Lalb_Chr04	16,724,499	Rychel-Bielska et al. 2020	published		field				
Lalb_Chr05	2,957,601	Alkemade et al. 2022	published	GWAS	controlled conditions				
Lalb_Chr18	12,358,001	Schwertfirm et al. 2024	published	GWAS	field				
Lalb_Chr21	3,907,702	Schwertfirm et al. 2024	published		field				
Lalb_Chr24	394,166	Schwertfirm et al. 2024	published		field				
Lalb_Chr04	2,315,230	Schwertfirm et al. 2024	*de novo*/published		field	1.39×10^{-6}	32	251/255	GLM
Lalb_Chr10	16,065,293	Schwertfirm et al. 2024	*de novo*/published		field	2.97×10^{-10}	61	255/255	GLM
Lalb_Chr06	2,629,006	Schwertfirm et al. 2024	*de novo*		field	6.12×10^{-7}	55	57/255	GLM
Lalb_Chr06	7,243,029	Schwertfirm et al. 2024	*de novo*		field	1.99×10^{-6}	65	47/255	GLM
Lalb_Chr06	7,349,762	Schwertfirm et al. 2024	*de novo*		field	1.99×10^{-6}	51	67/255	GLM
Lalb_Chr08	11,779,503	Schwertfirm et al. 2024	*de novo*		field	1.00×10^{-7}	45	254/255	GLM
Lalb_Chr10	15,808,902	Schwertfirm et al. 2024	*de novo*		field	4.88×10^{-8}	43	255/255	GLM

[a] – "*de novo*" stands for markers identified in the present study, while "published" characterize markers from literature

New markers from genomic prediction re-analysis of Alkemade et al. (2022) dataset

Genomic prediction was newly conducted on the Alkemade et al. (2022) dataset (raw data as provided in the original study) in R version 4.2.1 (R Core Team, 2022) using *rrBLUP* version 4.6.1 (Endelman 2011).

Missing genotypic information was imputed using the function A.mat with mean imputation method and a maximum missing marker information of 0.5. Genomic prediction was tested in a cross validation set-up of combinations of different (i) genotype numbers (ranging from 50 to 170), (ii) percent of training from 0.3 to 0.9 increasing by 0.1 increments; (iii) phenotypic values from controlled conditions presented in Alkemade et al. (2022), namely AUDPC and sAUDPC, running 20 iterations for each cross validation. The $(AUDPC)^2$ value was found to be most descriptive based on skewness, thus was selected as the variable for marker definition, but marker effects and predicted phenotypic values were re-calculated with the sAUDPC score, which is more consistent with literature.

After running the GP with all genotypic data included, several rounds of marker thinning to achieve a relevant, yet minimized SNP-marker panel, were performed as follows: first, initially running the GP and omitting markers below the absolute cut-off value with minimal effect sizes for consideration (being minimum/maximum marker effect multiplied by 0.6). The GP was repeated using the thinned marker list and its performance was compared to a randomly sampled marker set of the same marker density to identify most relevant markers. Following, additional rounds of thinning and comparisons were applied by dropping markers one at a time and rerunning the GP to check how the prediction accuracy changed. If the change in correlation was neutral or had a negative impact on the prediction accuracy, that marker was omitted from the marker set due to irrelevance, which thinning step was repeated three more times. Subsequently, by running ANOVA tests, *p*-values (*p*) were estimated for variations among the two allelic groups and explained variance values were calculated for each marker. Markers with $p < 0.005$ and explained variance $> 10\%$ were designated as highly relevant, and the best six were used to generate a genome-wide haplotype. Using these haplotypes, *p*- and explained variance values were estimated once more with haplotypes with $p < 0.005$ and explained variance $> 70\%$ designated as relevant. These six markers, with the remaining markers from the last marker thinning step were used for subsequent assay design and genomic prediction, selecting 50 SNP markers in total.

This resulted in 50 highly informative markers. However, six of them contained too many polymorphic loci in the context sequence to design PACE assays, and two additional more failed at the PACE assay reaction step due to ambiguous amplification confirmed by a melt step. These eight SNPs represented marker effects ranging between −0.0196 and 0.0202. Therefore, a final GP model, based on the technically working 42 markers, (Table 3, Supplementary Table S7) was used in our study and compared for performance in the training and validation sets.

Table 3 Distribution of 42 SNP markers used for genomic prediction across the 25 chromosomes of white lupin. Detailed information on the loci is presented in Supplementary Table S7

Chromosome	Marker number
Lalb_Chr01	4
Lalb_Chr02	5
Lalb_Chr04	4
Lalb_Chr05	1
Lalb_Chr06	3
Lalb_Chr10	4
Lalb_Chr11	2
Lalb_Chr13	1
Lalb_Chr14	3
Lalb_Chr15	2
Lalb_Ch17	4
Lalb_Chr18	4
Lalb_Chr19	1
Lalb_Chr20	1
Lalb_Chr21	1
Lalb_Chr22	1
Lalb_Chr23	1

Genomic prediction model application and validation

42 loci (Table 3) were used for genomic prediction performance evaluation, in the three validation panels. The analysis, as for the model training, is based on a ridge regression, conducted in R version 4.4.2 (R Core Team, 2024) using rrBLUP version 4.6.3 (Endelman 2011). For validation using the original training dataset, missing allelic information was imputed using A.mat with the mean method with maximum missing marker information set to 0.1. The imputed allelic matrix was multiplied by the respective marker effect, and the values were adjusted by the beta, both predefined during model training using the training panel`s phenotypic values with the 42 successful SNP-marker genotyping data. The genomic prediction was performed with 20 k-fold cross-validations in the original training datasets. Due to sample size in the validation panels, 10 k-fold cross-validations were performed for the multi-parental panel (field phenotyping, $n=96$), and 5 k-fold cross-validation for the diversity- ($n=48$), multi-parental- (controlled conditions phenotyping, $n=22$) and biparental ($n=50$) panels. Mean and standard error of the correlation (Pearson correlation, r) between predicted and observed phenotypic values, root mean square deviation (RMSE, Sharma et al. 2022) and coefficient of determination (R^2, Wright 1921) values were calculated across cross-validations.

In addition, improved genomic prediction models included retraining with the different panels using the same cross-validation approach as described above. The retrained datasets consisted of data from one validation panel, merged together with the data from the initial training panel for a total of four re-trained models (Table 4). For datasets originating from controlled conditions, adjusted phenotypic scores were recalculated for the merged retraining dataset as described above, while for those from the field, the original adjusted phenotypic values from the training panel were merged with the field phenotypic values of the respective dataset.

The allelic states of the 42 SNP markers were also used for a principal component analysis (PCA) to highlight intra- and interpopulation structure, created using *PCAtools* version 2.14.0 (Blighe and Lun 2023).

Results

Anthracnose resistance in diversity, multi-parental and biparental panels

No accession showed full resistance to anthracnose, neither in the climate chamber, nor in the field. Both diversity- and multi-parental panels showed clear differentiation between highly susceptible accessions and most resistant ones, with a continuous range of values for the accessions in between (Supplementary Table S1, Supplementary Table S3).

The diversity panel (Fig. 1A, yellow) was curated to represent the span of resistance status of the global genepool, but focusing on enriching resistant accessions, which are scarce. The adjusted sAUDPC scores ranged from 2.65 (± 0.28) to 5.37 (± 0.28), with three breeding lines from the FiBL-gzpk breeding programme surpassing the most resistant variety on the market, `Frieda` (Supplementary Table S1). In the multi-parental panel`s phenotypic data from controlled conditions (Fig. 1A, reddish brown), the values ranged from 3.84 (± 0.30) to 5.29 (± 0.23) (Supplementary Table S3). The cultivar `Frieda` performed worse than most of the breeding lines (sAUDPC score = 4.67 ± 0.22), and `Amiga` performed the worst out of all included accessions. Conversely, in the field (Supplementary Table S2, Fig. 1B, reddish brown), the multi-parental panel ranged between 4.05 and 7.85 (sAUDPC scores evaluated for individual plants in this case), and `Frieda` performed the third best (sAUDPC score = 4.25). The F_3 individual plants in the biparental panel (Supplementary Table S4, Fig. 1B, blue) showed great differentiation in anthracnose (single late-season disease score) values ranging from 4 to 9.

Within our multi-parental panel, comparing the field and controlled conditions observations, the phenotype of individuals in the field and the progeny of 22 of the exact same plants under controlled conditions did not correlate ($p = 0.718$). However, the correlation value between controlled conditions and field phenotyping (plot level) in a subset of 12 accessions from the training panel used for the genomic prediction where both scorings were available was 0.75 ($p = 0.004$).

Broad sense heritability values for the experiments from controlled conditions (where biological replication was available) were moderate to high, 0.70 for the training panel, 0.78 for the diversity- and 0.57 for the multi-parental panel (Fig. 1A).

Testing marker-trait associations from literature in diversity- and breeding panels

Of the three markers reported by Rychel-Bielska et al. (2020) (Table 2), both Lalb_Chr02_550648 and Lalb_Chr04_16724499 were present in the raw SNP dataset reported by Schwertfirm et al. (2024); however, only Lalb_Chr02_550648 was significantly associated with anthracnose score in that dataset ($p = 1 \times 10^{-7}$). Lalb_Chr02_550648 also showed significant association with the phenotype in the

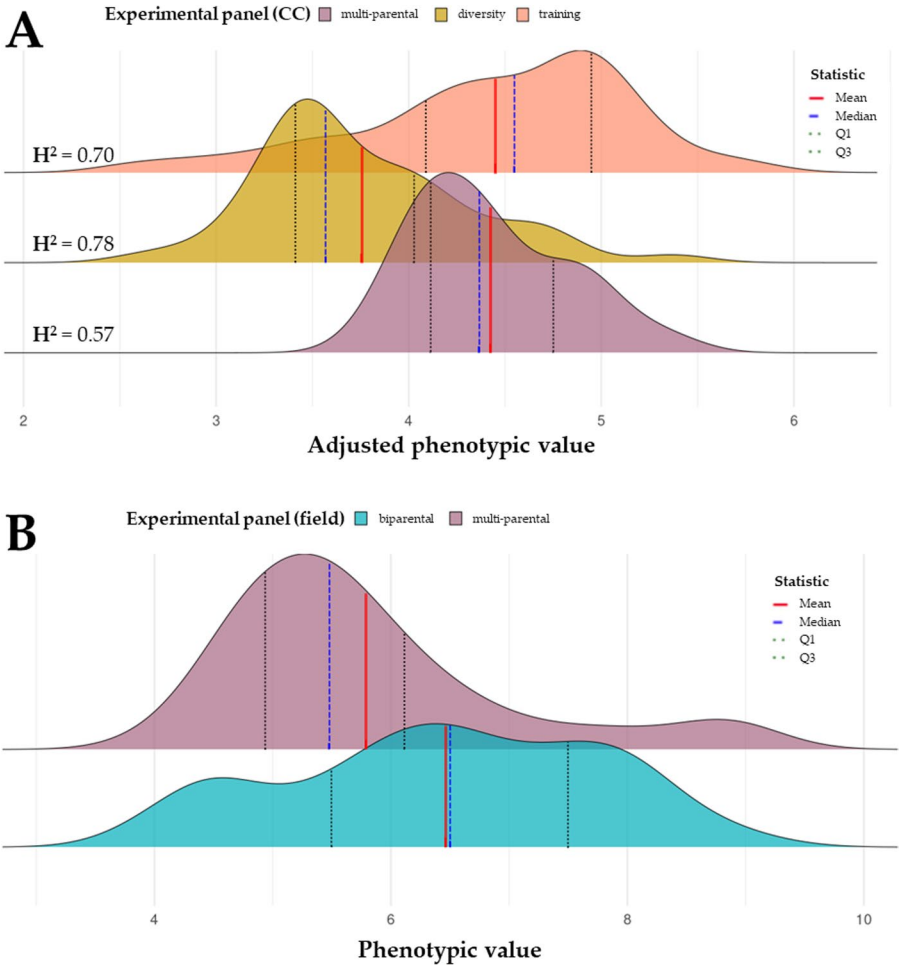


Fig. 1 Ridge plots of the distribution of phenotypic data of the validation panels, marking mean (red line), median (blue dashed line), Q1 and Q3 (black dotted lines) values in each dataset. (**A**) Ridge plots of the adjusted sAUDPC values (6 biological replicates) originating from climate chamber trials (**B**) Ridge plots of the phenotypic values (single plant observations) originating from field observations, sAUDPC values for the multi-parental panel (reddish brown) and a single late-season disease score for the biparental pane (blue). The broad sense heritability (H^2) value is marked in each sub-plot

multi-parental panel based on field observations ($p=0.01$; Fig. 2D; Supplementary Table S7). The marker Lalb_Chr02_635857 was monomorphic in all datasets except the diversity-panel, where it showed no significant association with the phenotype ($p=0.85$).

Allelic groups at Lalb_Chr05_2957601, a locus selected from the three reported by Alkemade et al. (2022), were not associated with differences in anthracnose severity in any of the three evaluated panels (Fig. 2C).

The performance of five high-confidence SNP markers identified by Schwertfirm et al. (2024) was further assessed across all validation panels and compared with

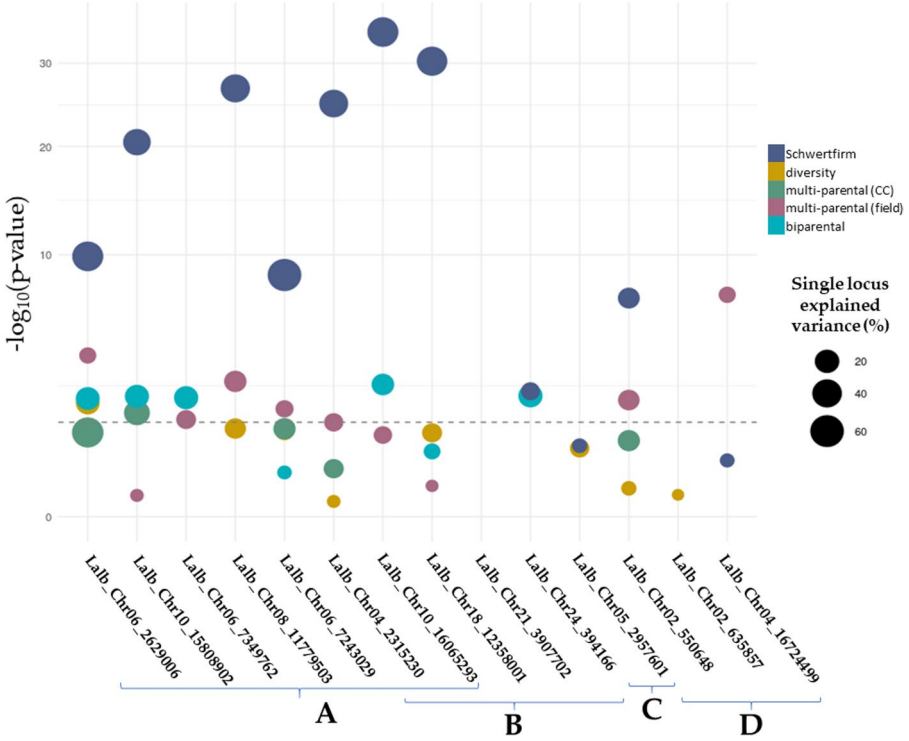


Fig. 2 Scatter plot scaled to size of the proportion of variance explained of SNP-markers from literature and newly identified. Marker names on the x-axis refer to the individual SNPs and, on the y-axis their respective -$\log_{10}$(*p*-value) from analysis of variance of phenotypic value across the allelic states, with significance threshold of $p = 0.05$ (dotted line). Phenotypic variance explained (PVE) values were calculated for the validation panels based on all individual SNP markers. (**A**) 8 markers originating from the present study, using the raw data from Schwertfirm et al. (2024). (**B**) 5 high confidence SNP-markers originating from Schwertfirm et al. (2024) with successful PACE-assays out of the 7 targeted. (**C**) 1 marker originating from Alkemade et al. (2022). (**D**) 3 markers originating from Rychel-Bielska et al. (2020). Colours show which of the evaluation panels the dot refers to" (navy – original dataset presented in Schwertfirm et al. 2024; yellow – diversity panel; green – multi-parental panel using phenotypic data from the climate chamber; reddish brown – multi-parental panel using phenotypic data from the field; light blue – biparental panel), and the size refers to the explained variance value

their performance in the original dataset. Across all validation panels, we observed a consistent reduction in explained variance and strength of association relative to the original results (Fig. 2B). Regarding polymorphism, three markers were polymorphic in the diversity-panel, two in the multi-parental panel under climate chamber conditions, three in the multi-parental panel based on field data, and four in the biparental panel. In terms of association with anthracnose severity, none of the markers showed significant associations in either the diversity-panel or the multi-parental panels. In contrast, two of the four polymorphic SNPs in the biparental panel (Lalb_Chr24_394166 and Lalb_Chr10_16065293) were significantly associated with disease score at a threshold of $p = 0.05$ (Fig. 2B).

Two loci overlapped between the original GWAS reported by Schwertfirm et al. (2024) and the GWAS conducted in the present study on the same dataset (Lalb_Chr04_2315230, Lalb_Chr10_16065293). Of these, none showed significant associations with anthracnose severity in the validation populations, except Lalb_Chr10_16065293, which was significantly associated in the biparental panel ($p = 0.005$; Fig. 2A–B overlap; Supplementary Table S7).

Using all genotypic data of individually assessed loci together to calculate the total phenotypic variance explained (Fig. 2), the model explains 21% and 24% of the observed phenotypic variance in the field panels. For the diversity panel, although the trait shows high broad-sense heritability ($H^2 = 0.78$, Fig. 1A), the fitted single marker-based model using the individually assessed markers together, captures only a modest proportion of total phenotypic variance (PVE = 15%). For the multi-parental panel, moderate-to-high heritability ($H^2 = 0.57$, Fig. 1A) comes with a higher PVE (PVE = 26%).

Testing newly identified marker-trait associations in diversity- and breeding panels

Apart from the two aforementioned loci overlapping with the original publication, this study identified five additional SNP-markers (Fig. 2A), originating from the re-analysis of the dataset published in Schwertfirm et al. (2024). This re-analysis consisted of a complementary GWAS to the original dataset presented.

In the validation, no single marker was consistently associated to the anthracnose trait within all the investigated panels. However, Lalb_Chr06_2629006 was significantly associated to the disease phenotype in the diversity-, multi-parental- (field phenotyping) and the biparental panel with single locus explained variance values of 0.18, 0.03 and 0.18 respectively, with clear differences between different allelic groups in the latter two panels (Fig. 3). One additional SNP-marker Lalb_Chr10_15808902 was significant in the multi-parental (climate chamber phenotyping) and the biparental panels, but association of the allelic state was inconsistent across the two cases. Three additional SNP-markers from the five newly identified loci had a significant association to the anthracnose phenotype in at least one validation panel. Two located on chromosome *Lalb_Chr06* (Lalb_Chr06_7349762, Lalb_Chr06_7243029), and one on chromosome *Lalb_Chr08* (Lalb_Chr08_11779503), with significant associations with field observations (multi-parental and biparental panel, Fig. 3).

Testing accuracy of genomic prediction model based on a minimal number of markers

As for diversity, compared to the 42 markers selected for genomic prediction in the training panel (original dataset from Alkemade et al. 2022), all are polymorphic in the diversity-panel, while 39 and only 16 are polymorphic in the multi-parental and in biparental panel, respectively.

The first two components of the PCA (Fig. 4), explaining 17.28% and 9.14% of variation using all 42 markers (PC03-37 ranging from 6.1 to 0.3%), reveal that the training panel is closer related to the diversity-panel compared than to the multi-

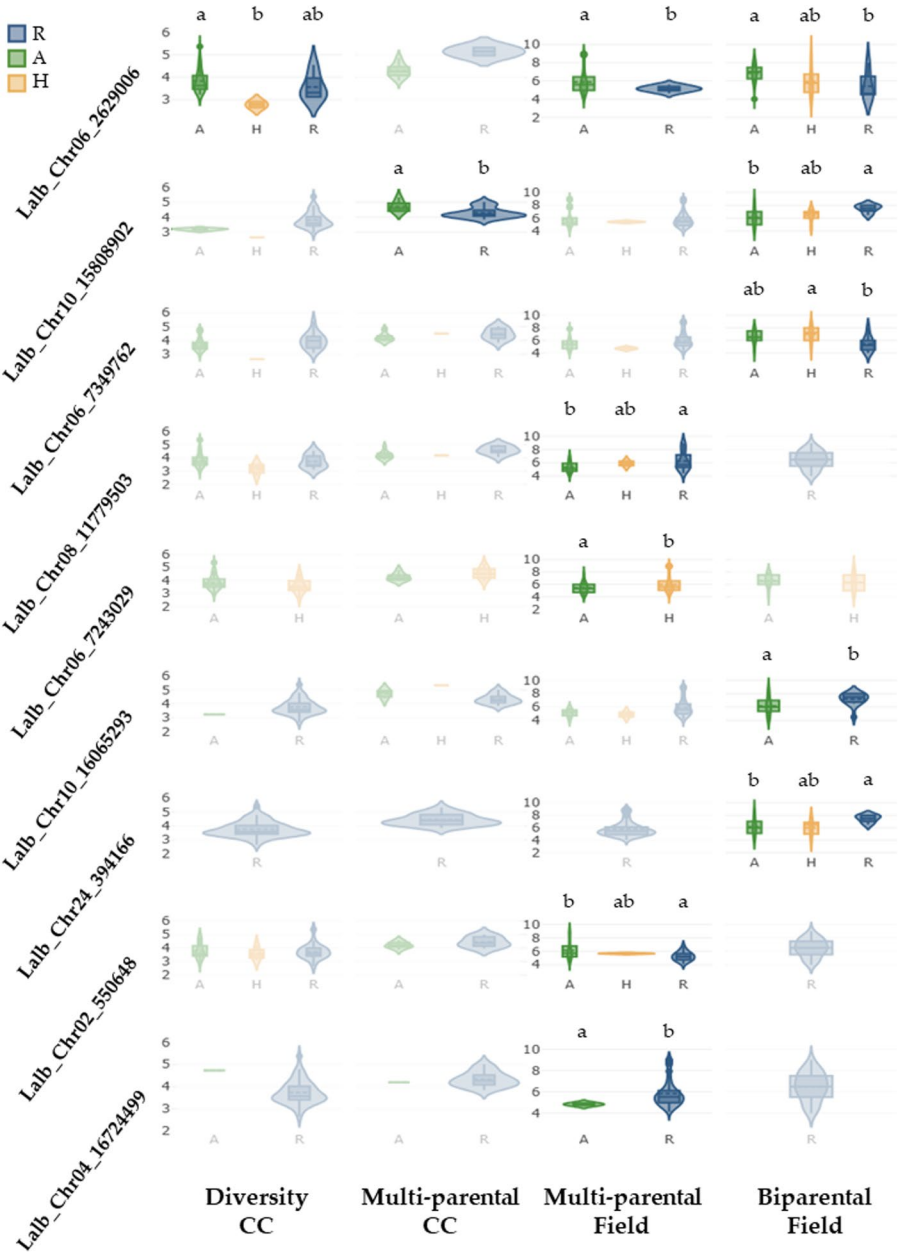


Fig. 3 Violin plots of different allelic states and their associated phenotypic values (sAUDPC on the y-axis) with statistical grouping in case of significant association in at least one validation panel. The colour blue refers to individuals carrying the reference allele (R, allelic state associated to the reference genome of the cultivar `Amiga`), green to the individuals carrying the alternative allele (A), and yellow to heterozygous individuals (H). Blurred windows show no significant variations between the alleles in regard to the anthracnose level. Letters above the clear window allele comparisons indicate significance groups ($p < 0.05$)

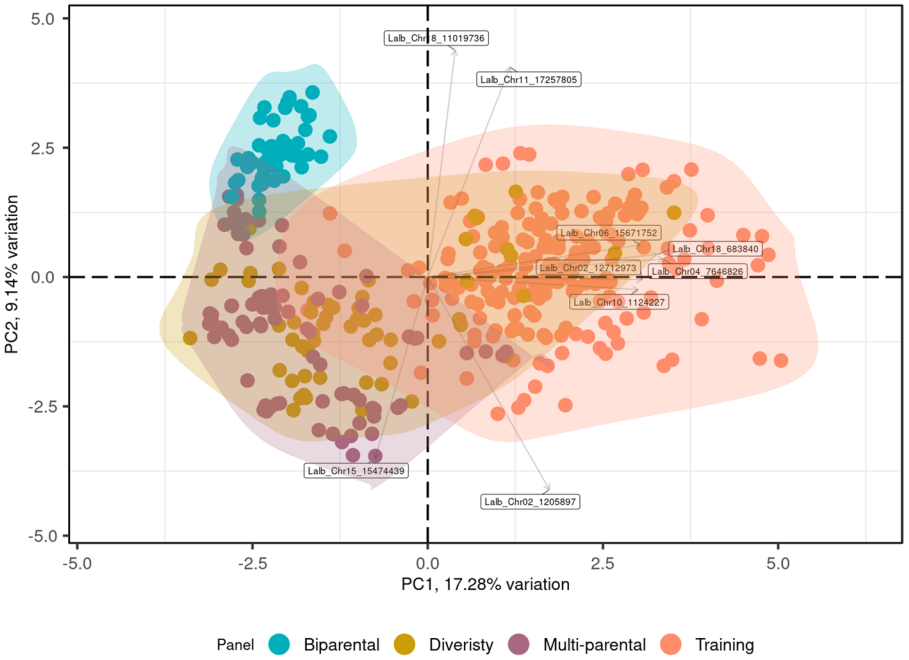


Fig. 4 First two principal components (PC) of PCA representing all datasets used in the study, colours, shapes and encircling indicate different panels. All SNP-markers ($n=42$) included in the analysis

parental panel. The individuals of the multi-parental panel, clustering together with those of the training panel, represent three independent breeding lines carrying resistance sources from Chilean breeding programme and Ethiopian genetic resources, which resistance sources were present in the training panel as well (Fig. 4). Noticeably individuals of the biparental panel were clustering most distantly from the training panel.

The genomic prediction based on the 42 markers set had a prediction accuracy (r) of 0.84 ± 0.13 across 20 cross-validations in the training panel. The coefficient of determination ($R^2 = 0.58 \pm 0.34$) also signalled that the model accurately predicts the phenotypes, although with high levels of variation amongst cross-validations (Fig. 5A; Table 4). In the diversity-panel (42 markers), the correlation between predicted and observed phenotypic values across 5-fold cross-validations had a prediction accuracy of $r = 0.30$, with high standard error across cross-validations ($SE = 0.33$). Although the RMSE value (0.95 ± 0.07) was within acceptable range, the coefficient of determination $R^2 = -2.89 \pm 1.89$ highlights the failure of the initial model (Fig. 5B; Table 4). There is no significant correlation between predicted and observed values for the multi-parental panel (39 polymorphic loci, Fig. 5C-D) and the biparental panel (16 polymorphic loci, Fig. 5E; Table 4).

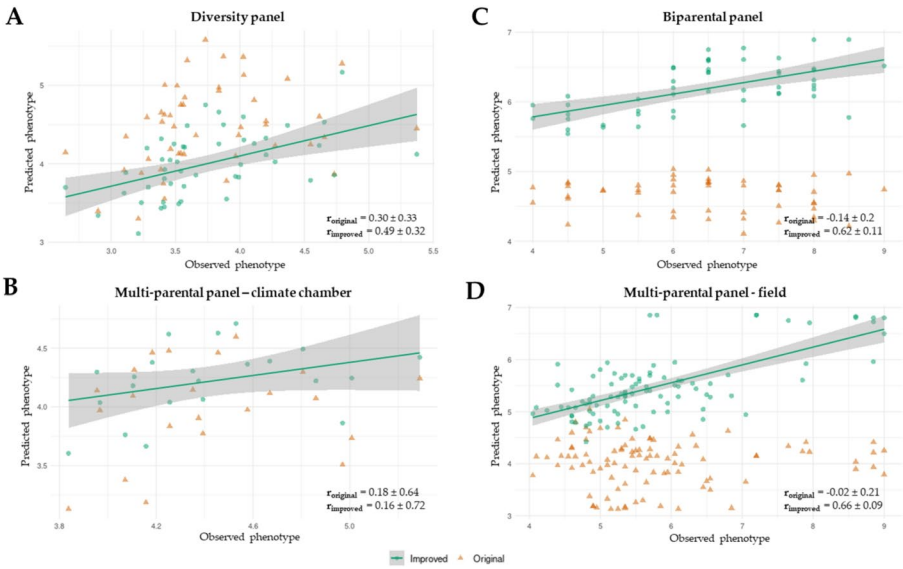


Fig. 5 Scatter plots of predicted and observed phenotypic values, presenting original ($r_{original}$) and improved ($r_{improved}$) prediction accuracy. (**A**) Scatter plot of predicted and observed phenotypic values in the diversity panel ($n=48$), using 42 polymorphic SNP-markers. (**B**) Scatter plot of predicted and observed phenotypic values in the multi-parental panel using the phenotypic value from controlled conditions ($n=22$), using 39 polymorphic SNP-markers. (**C**) Scatter plot of predicted and observed phenotypic values in the biparental panel using the phenotypic value from the field ($n=50$), using 16 polymorphic SNP-markers. (**D**) Scatter plot of predicted and observed phenotypic values in the multi-parental panel using the phenotypic value from the field ($n=96$), using 39 polymorphic SNP-markers

Retraining genomic prediction model enhances prediction accuracy

As the initial estimated genomic prediction model was unsuitable to make predictions beyond the genomic diversity boundaries it was trained on, we aimed to improve the genomic prediction model. Therefore, we made use of the additional source of diversity found in the diversity-, multi-parental and biparental panels, keeping the same 42 markers (Table 4). For GP1 and GP2 adjusted sAUDPC values were used, while for GP3 and GP4, phenotypic values from the climate chamber and the sAUDPC values from the field observations were merged before the model training.

Genomic prediction performance was evaluated by comparing the original model trained on the training panel alone (Fig. 5; Table 4, GP0) with models retrained using merged datasets (Fig. 5; Table 4, GP1–GP4). Based on 20-fold cross-validation within the respective training datasets, the highest prediction accuracy was obtained using the training panel only ($r=0.84\pm0.13$). All retrained models showed stable predictive performance, albeit with slightly reduced accuracy (Fig. 5; Table 4, GP1–GP4).

When applied to the original training dataset, retrained models exhibited comparable performance, with similarly low mean RMSE values and comparable coefficients of determination. Models retrained with datasets derived from controlled phenotyping conditions (Fig. 5A-B; Table 4, GP1–GP2) showed reduced standard errors across cross-validations, indicating increased prediction stability. In contrast, models

Table 4 Prediction accuracy (r), RMSE and R^2 values and respective standard error of retrained GP models (GP1-GP4) and of the original model (GP0), as well as the aforementioned metrics for genomic prediction in all panels included in the study. Each column represents a different combination of validation panels used for model training together with the training panel. 20-fold cross validations were performed for all training datasets, 10-fold for validation in the multi-parental panel (field observations), and 5-fold for the remaining models. Results of the respective dataset used for re-training in black, other combinations in dark grey

			GP0	GP1	GP2	GP3	GP4
Retraining Model improvement		n	192	240	214	288	242
		Training panel	X	X	X	X	X
		Diversity-panel		X			
		Multi-parental panel (CC)			X		
		Multi-parental panel (Field)				X	
		Biparental panel (Field)					X
		Prediction Accuracy	0.84 ± 0.13	0.78 ± 0.09	0.82 ± 0.09	0.70 ± 0.18	0.77 ± 0.19
		RMSE	0.37 ± 0.08	0.45 ± 0.05	0.37 ± 0.13	0.73 ± 0.16	0.68 ± 0.16
		R^2	0.58 ± 0.34	0.58 ± 0.14	0.60 ± 0.18	0.39 ± 0.28	0.53 ± 0.32
Application Genomic prediction	Training panel	Prediction Accuracy		0.79 ± 0.14	0.83 ± 0.13	0.56 ± 0.23	0.70 ± 0.19
		RMSE		0.43 ± 0.09	0.38 ± 0.09	0.61 ± 0.16	0.50 ± 0.12
		R^2		0.42 ± 0.47	0.56 ± 0.34	-0.10 ± 0.82	0.10 ± 1.01
	Diversity panel	Prediction Accuracy	0.30 ± 0.33	0.49 ± 0.32	0.29 ± 0.32	-0.002 ± 0.40	0.04 ± 0.38
		RMSE	0.95 ± 0.07	0.54 ± 0.07	0.98 ± 0.07	1.60 ± 0.09	1.22 ± 0.18
		R^2	-2.89 ± 1.89	-0.26 ± 0.71	-3.10 ± 1.91	-9.31 ± 3.17	-5.30 ± 3.17
	Multi-parental panel (CC)	Prediction Accuracy	0.18 ± 0.64	0.16 ± 0.72	0.31 ± 0.60	-0.10 ± 0.60	0.11 ± 0.58
		RMSE	0.62 ± 0.23	0.77 ± 0.22	0.41 ± 0.17	0.94 ± 0.23	0.48 ± 0.15
		R^2	-3.86 ± 3.0	-6.97 ± 4.57	-1.07 ± 1.27	-12.01 ± 9.91	-2.17 ± 1.88
	Multi-parental panel (Field)	Prediction Accuracy	-0.02 ± 0.21	- 0.02 + 0.21	0.01 ± 0.22	0.66 ± 0.09	0.01 ± 0.21
		RMSE	2.15 ± 0.34	2.37 ± 0.34	1.94 ± 0.33	0.92 ± 0.22	1.86 ± 0.33
		R^2	-3.08 ± 1.60	-3.95 ± 1.96	-2.25 ± 1.10	0.32 ± 0.08	-1.99 ± 1.00
	Biparental panel	Prediction Accuracy	-0.14 ± 0.2	-0.22 ± 0.21	-0.12 ± 0.22	-0.05 ± 0.19	0.62 ± 0.11
		RMSE	2.24 ± 0.35	2.48 ± 0.33	2.12 ± 0.35	1.82 ± 0.34	1.12 ± 0.32
		R^2	-2.5 ± 1.44	-3.31 ± 1.74	-2.15 ± 1.26	-1.27 ± 0.82	0.20 ± 0.13

retrained using datasets that included field-phenotyped panels (Fig. 5C-D; Table 4, GP3–GP4) displayed increased mean RMSE values and greater variability, consistent with the higher phenotypic variance observed under field conditions (Fig. 1B). The coefficient of determination remained high when the training panel was merged with the biparental panel (Table 4, GP4), whereas a reduction in R^2 was observed when merged with the multi-parental panel field phenotypes (Table 4, GP3).

Retraining using field-derived phenotypic datasets resulted in effective prediction within their respective validation panels. For the multi-parental panel, prediction accuracy reached $r_{improved} = 0.66 \pm 0.09$ with a coefficient of determination of $R^2 = 0.32 \pm 0.08$ (Fig. 5D; Table 4, GP3). The corresponding mean RMSE (0.92 ± 0.22) reflected the broader phenotypic range present in field observations. Similarly, the model retrained with the biparental panel exhibited a prediction accuracy of $r = 0.62 \pm 0.10$ and $R^2 = 0.20 \pm 0.13$, accompanied by higher RMSE values (1.12 ± 0.32), consistent with increased phenotypic dispersion (Fig. 5C; Table 4, GP4).

In contrast, retraining did not consistently improve predictive performance for validation panels with phenotypes originating from controlled conditions. The model retrained with the diversity panel (Fig. 5A; Table 4, GP2) showed substantial variability across cross-validations ($r = 0.49 \pm 0.32$) and limited predictive ability. Although mean RMSE values remained stable (0.54 ± 0.07), coefficients of determination were highly variable and frequently negative ($R^2 = -0.26 \pm 0.71$), indicating poor model transferability in this scenario.

Discussion

Our analysis of the performance of anthracnose resistance markers in the context of a small- scale white lupin breeding programme, aimed to close the gap between advanced molecular marker research and practical implementation of the results in plant breeding, while highlighting the challenges involved. For easy application in molecular marker-assisted selection, we worked with the cost-effective PACE genotyping, suitable also for small sized breeding programmes such as the organic one in Switzerland by FiBL and gzpk (a non-profit plant breeding association with more than 30 years of experience in organic plant breeding).

Our validation panels consisted of two breeding panels from the aforementioned breeding programme, and a diversity panel. The diversity-panel was curated to represent the span of the phenotypic variability of the trait, maximising the presence of phenotypically resistant accessions. This was meant to complement and validate the findings of the GWAS studies presented on the two panels (LUW and training) presented by Alkemade et al. (2022) and Schwertfirm et al. (2024). Comparing the broad sense heritability (H^2), we observed (Fig. 1A) high values for the training and diversity-panels (0.70 and 0.78, respectively). However, in the diversity-panel, only a limited fraction of the total phenotypic variance (PVE = 15%) is captured by the predictors included in the model, highlighting the limitation of genetic architecture of training vs. validation panels. This limitation is also observed in the multi-parental panel using the climate chamber values ($H^2 = 0.57$, PVE = 26%).

Individually assessing SNP loci described in previous publications, no single marker from literature was consistently successful in discriminating the anthracnose disease level across all panels used for validation. The most significant locus from the study by Schwertfirm et al. (2024), Lalb_Chr10_16065293 (-log(p) = 14.91), was significant in the biparental panel, where one of the parents is the cultivar `Frieda`. This is particularly interesting since 41 out of 73 breeding lines from the panel of the original study clustered together with the cultivar `Frieda`, leading to the assumption that the resistance source present in `Frieda` is overrepresented in this panel, thus the marker-trait association presented there would also contain the same source of resistance. This specific SNP on chromosome *Lalb_Chr10* is also in the QTL region that the latest white lupin linkage map presents for anthracnose resistance (Książkiewicz et al. 2017), a finding consistent with ca. ten years of research. However, it is not located in the coding sequence of any gene, rather within a mRNA cluster (cluster_56277), which are known epigenetic regulatory molecules potentially associated to plant defence (Yang et al. 2021; Luo et al. 2024; Mierziak and Wojtasik 2024). We found no paralogs or orthologs through nBLASTs (Altschul et al. 1990; Boratyn et al. 2012; Camacho et al. 2009; Morgulis et al. 2008; Zhang et al. 2000), thus no further information could be concluded regarding its function or involvement in defence against fungal pathogens.

As a second step, this study aimed to expand the utility of already existing datasets with the re-analysis for further marker discovery through GWAS in the dataset Schwertfirm et al. 2024 and through GP in the dataset Alkemade et al. (2022).

Schwertfirm et al. (2024) present a GWAS based on multi-locus models (MLMM, FarmCPU and BLINK) applied to the pre-filtered marker set for missing data and minimum allele frequency. The re-analysis in our study aimed to focus on SNPs which were lost during the filtering steps done in the original publication. GLM enabled the inclusion of these markers, preserving potentially informative genetic variation (Kaler et al. 2020). The approach is considered robust as the performance of all newly identified markers was assessed in all environments and populations included for adequate validation.

All five additional SNP-markers from the repeated GWAS are significant in at least one validation panel, with one marker (Lalb_Chr06_2629006) being significant in three out of the four different validation panels highlighting the specificity based on the genetic origin of the resistance.

To exploit a different source of markers and selection approach, we performed genomic prediction in the data set by Alkemade et al. (2022). We developed a model based on a limited number of 42 markers as blue-print for low-cost application of genomic selection in small scale breeding programmes for hard to phenotype, polygenic traits. There was no possibility to cross-validate findings from the two derived SNP datasets (Alkemade et al. 2022; Schwertfirm et al. 2024) due to the poor overlap between raw SNPs, as only 14 loci within the 42 marker panel could be found in the dataset of Schwertfirm et al. (2024).

Compared with the high prediction accuracy of the original genomic prediction (GP) model in the training panel ($r_{original}$ = 0.84 ± 0.13), predictive performance dropped substantially in all independent validation panels. In these panels, correlations between predicted and observed anthracnose scores were close to zero, and in

several cases the coefficient of determination (R^2) was negative, indicating poor and unstable prediction. A major limitation affecting model transferability is the reduced allelic diversity in the validation panels relative to the training population. Of the 42 SNP markers included in the original model, only 39 were polymorphic in the multi-parental panel and only 16 in the biparental panel. This reduction reflects the narrower genetic base of the white lupin breeding panels and limits the model's ability to capture the full range of marker effects identified during training. Recalculating marker effects using only the polymorphic markers (39 or 16 SNPs, respectively) did not improve predictive performance, suggesting that the loss of informative loci—and not simply incorrect estimation of marker effects—contributes to the observed decline in accuracy. Even in the diversity-panel, which was genetically most similar to the training panel (Fig. 4) and in which all 42 markers were polymorphic, prediction accuracy remained highly variable ($r_{original} = 0.30 \pm 0.33$) with coefficient of determination values consistently negative ($R^2 = -2.89 \pm 1.89$). This instability may be connected to the enrichment of resistance sources, determined by phenotyping results rather than genotyping data in the diversity-panel compared to the training panel. Differences in sample size, trait distribution, and phenotypic variance between the training and validation datasets likely further reduced model performance. In addition, mismatches between phenotyping environments may have contributed to the poor transferability of the GP model. Validation panels included data from controlled environments and field conditions that differed from those used for model training (controlled conditions only), potentially altering the expression of quantitative resistance and the relevance of specific marker effects. Marker–environment interactions may therefore play an important role in limiting model stability. Overall, the consistently low predictive ability observed across validation panels (Table 4, GP0) highlights the combined impact of limited panel diversity, environmental heterogeneity, and marker–environment interactions on GP model robustness.

Most GP models use a much higher marker density, although the increase in available marker information does not always correlate with increased prediction accuracy (Fu et al. 2017; Zhao et al. 2025). In a recent study about genomic prediction in white lupin (Annicchiarico et al. 2025), the authors present a genomic selection model for key agronomic traits of seed weight, protein- and oil content as well as total quinolizidine alkaloid content focusing on intra- and inter-population, as well as intra- and inter-environment scenarios. Despite of using over 30,000 SNPs for prediction in all scenarios, low prediction accuracy scores were observed in inter-population and inter-environment scenarios when comparing breeding lines to landraces, consistent with our findings.

From another pathosystem, Juliana et al. (2017) report analyses for leaf-, stem- and stripe rust resistance in wheat, using 5,102 and 8,066 marker loci for the prediction, respectively. Even with the magnitudes higher marker density than presented in this study, the prediction accuracy was 0.39 when working with one wheat dataset. A different study on wheat focused on *Fusarium* head blight for which the resistance (similarly to anthracnose resistance in white lupin) is polygenically controlled. Using 4,205 SNPs for the genomic prediction, the authors report prediction accuracies ranging between 0.4 and 0.6 for disease severity and -incidence using *rrBLUP* (Zhang et

al. 2021). In the present study, four sequential marker thinning steps were included to reach 42 highly informative markers, and the prediction accuracy dropped substantially from the training to the diversity panel. In a study made on Norway spruce, the authors test the limitations of genomic prediction models by decreasing marker densities looking at different phenological characteristics. For bud burst, they report a prediction accuracy of 0.63 using 4,000 SNPs, which number decreased to 0.36 when using the allelic information of only 25 SNPs (Chen et al. 2023), which phenomenon was also reported in Cerioli et al. (2022) when using genomic prediction with systematically decreasing marker densities for agronomic traits in rice. Although, using a larger marker density does not per se correlate to more accurate predictions. For example, in a genomic prediction study for soybean mosaic virus resistance in soybean, researchers identified more accurate prediction accuracy using the top 100 loci, compared to the more than 8,000 available SNPs. The inclusion of all genotypic data available could create unwanted background noise (Zhao et al. 2025), potentially improved by continuous marker thinning steps. In particular, the approach used in this study (ridge regression), rely on effect shrinkage toward 0. Shrinking marker effects for predicting the phenotype of interest is the core strategy to limit false predictions. The more markers unrelated to the phenotype of interest are implemented in such models, the more shrinkage of marker effects needs to be performed to limit errors – resulting in suboptimal models. Other approaches, likes BayesB or lasso approaches might allow to reduce marker effects to 0 instead of a very small value – nevertheless, these models in general do not lead to drastically different prediction accuracies (Meuwissen et al. 2009; Puthiyedth et al. 2025). Therefore, “feature”, or marker selection, remains a robust approach to improve genomic prediction models. Although the minimal marker set showed to be successful in the training panel, it lacked predictive power in the validation panels, regardless of genetic background or panel size. This is not unexpected, especially considering that genomic prediction studies most often use subset-, interset- or progeny validation (Alemu et al. 2024; Crossa et al. 2025; Sallam et al. 2015) and higher marker density for complex traits, particularly in species with low linkage disequilibrium such as white lupin (Alemu et al. 2024). Increasing the marker quantity from 42 to 100 might have resulted in improved predictions but it was beyond the scope of this study. Another approach to improve the usefulness of GP is to extend the genetic diversity.

Genomic prediction models were retrained with the addition of a specific validation panel to the data from the training panel (Table 4, GP1-4), to improve the genomic prediction both from the facets of relatedness and sample size. Although still unable to accurately predict amongst validation panels, the retraining accentuates the possibility to still apply genomic prediction within unrelated material, with the inclusion of the panel of accessions to be validated into the genomic prediction model. The biggest limitation in our study was size of the specific validation panel. The least successful retrained model when applied to its corresponding validation panels was with the multi-parental panel, using the phenotypic values from controlled conditions ($n=22$), with the highest deviation in across the cross validations and minimal success even in the best scenario (Fig. 5B; Table 4, GP2). Moreover, narrower genetic diversity within both breeding populations and the diversity panel highlights the limi-

tations. When retraining the model with the diversity panel (Fig. 5A; Table 4, GP1) included in the model training, the deviation remains high in the prediction accuracy ($r = 0.49 \pm 0.32$), and the coefficient of determination ($R^2 = -0.26 \pm 0.71$), highlighting great differences between cross-validations. The barrier that the retraining overcame is applying genomic prediction models trained with phenotypic values originating from controlled conditions for validation in panels where observations were made on the field. The combination of training panel and the multi-parental panel (field observations) showed the greatest improvement (Table 4, GP3), with reliable prediction accuracy ($r = 0.66 \pm 0.09$) and coefficient of determination ($R^2 = 0.32 \pm 0.08$) across cross-validations, where the moderately high RMSE values (0.92 ± 0.22) could be explained by greater differentiation in phenotypic values observed on the field. This successful application, to a lesser extent, is also highlighted by the results of the model retraining with the biparental panel (Table 4, GP4). This scenario accentuates successful application in breeding panels with low genetic diversity, such as an F_3 panel originating from a biparental crossing, regardless of the differences in phenotyping approaches (sAUDPC vs. single disease score). Thus, to address the inconsistencies between observations made on the field and in controlled conditions, the best suggested approach would be using the combination of marker loci presented in Schwertfirm et al. (2024), which study worked with phenotypes coming from the field, and those coming from Alkemade et al. (2022), which included phenotypic evaluation under controlled conditions, and taking into account marker-environment interactions, as suggested by Skøt et al. (2024). With the combined approach using a selection of SNPs, the marker set would address environmental context-dependency intrinsically. Subsequently, using these loci for continuous selection would enable a more reliable genetic gain throughout the years.

Conclusion

With an in-depth analysis of previously published and newly detected loci associated to anthracnose resistance in white lupin, we provide a thorough investigation on the efficiency of currently available markers for marker-assisted selection. We confirmed SNP-markers associated to the previously reported QTL on chromosome *Lalb_10*, tagging the source of resistance associated to the cultivar 'Frieda'.

We also constructed a genomic prediction model based on a minimal number of 42 PACE markers for use in small-sized breeding programmes, partially positively validated in a diversity-panel, while highlighting the limitations in application when working with a narrower and unrelated genepool. The effect of training and validation set diversity, phenotyping environment, and marker–environment interactions on model stability are discussed against most recent GP literature. Our results show that a retraining of genomic prediction models leads to improved genomic prediction accuracy for anthracnose resistance, despite relatedness and different phenotyping platforms of validation panels, compared to the initial training dataset.

Abbreviations

CAPS	Cleaved Amplified Polymorphic Sequence
FiBL	Forschungsinstitut für Biologischen Landbau (Research Institute of Organic Agriculture)
GP	Genomic Prediction
GWAS	Genome Wide Association Study
gzpk	Getreidezüchtung Peter Kunz (Cereal Breeding Peter Kunz)
LfL	Bavarian State Research Center for Agriculture
LMM	Linear Mixed Model
MAS	Marker Assisted Selection
PACE	PCR Allelic Competitive Extension
PCR	Polymerase Chain Reaction
PVE	Phenotypic Variance Explained
RIL	Recombinant Inbred Line
rrBLUP	Ridge Regression Best Linear Unbiased Predictor
sAUDPC	Standard Area Under Disease Progress Curve
SNP	Single Nucleotide Polymorphism

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11032-026-01661-w.

Acknowledgements The authors wish to express their sincere gratitude to the farmers of Strickhof 339 (5324, Full-Reuenthal, Switzerland) for providing space for the field experiments, and to the temporary members of the FiBL Plant Breeding Team over the years for their valuable assistance with both field and laboratory work.

Author contributions **A. P.**: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing; **M. S.**: Conceptualization, Data curation, Formal Analysis, Investigation, Validation, Resources, Supervision, Writing – review & editing; **C. A.**: Investigation, Project administration, Funding acquisition, field trial conceptualization, Resources, Writing – review & editing; **M. K.**: Supervision, Writing – review & editing; **M. M. M.**: Conceptualization, Funding acquisition, Project administration; **G. S.**: Supervision, Writing – review & editing; **M. L.**: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.
All authors have read and approved the submitted manuscript, as well as any substantially revised version involving their contributions. Each author accepts responsibility for their own work and affirms their commitment to ensuring that any concerns regarding the accuracy or integrity of any part of the manuscript—even those in which they were not directly involved—are thoroughly investigated, appropriately resolved, and duly documented in the literature.

Funding This research has received funding from the projects LUPINNO SUISSE (funded by FOAG, grant agreement n. 2020/51/LUPINNO-SUISSE), LiveSeeding, (funded by the EU, SERI and UKRI, grant agreement n. 727230) and DIVINFOOD (funded by the EU, grant agreement n 101000383).

Data availability The authors used publicly available datasets available from supplementary material from Alkemade et al. (2022) and Schwertfirm et al. (2024). All data generated or analysed during this study are included in this published article and its supplementary information files.

Declarations

Competing interests The authors declare no competing interests.

References

Adhikari KN, Buirchell BJ, Thomas GJ, Sweetingham MW, Yang H (2009) Identification of anthracnose resistance in Lupinus albus L. and its transfer from landraces to modern cultivars. Crop Pasture Sci 60(5):472. https://doi.org/10.1071/CP08092

Adhikari KN, Thomas G, Diepeveen D, Trethowan R (2013) Overcoming the barriers of combining early flowering and anthracnose resistance in white lupin (Lupinus albus L.) for the Northern Agricultural Region of Western Australia. Crop Pasture Sci 64(9):914. https://doi.org/10.1071/CP13249

Alemu A, Åstrand J, Montesinos-López OA, Isidro y Sánchez J, Fernández-Gónzalez J, Tadesse W, Vetukuri RR, Carlsson AS, Ceplitis A, Crossa J, Ortiz R, Chawade A (2024) Genomic selection in plant breeding: Key factors shaping two decades of progress. Mol Plant 17(4):552–578. https://doi.org/10.1016/j.molp.2024.03.007

Alkemade JA, Messmer MM, Arncken C, Leska A, Annicchiarico P, Nazzicari N, Książkiewicz M, Voegele RT, Finckh MR, Hohmann P (2021a) A High-Throughput Phenotyping Tool to Identify Field-Relevant Anthracnose Resistance in White Lupin. Plant Dis 105(6):1719–1727. https://doi.org/10.1094/PDIS-07-20-1531-RE

Alkemade JA, Messmer MM, Voegele RT, Finckh MR, Hohmann P (2021b) Genetic diversity of Colletotrichum lupini and its virulence on white and Andean lupin. Sci Rep 11(1):13547. https://doi.org/10.1038/s41598-021-92953-y

Alkemade JA, Nazzicari N, Messmer MM, Annicchiarico P, Ferrari B, Voegele RT, Finckh MR, Arncken C, Hohmann P (2022) Genome-wide association study reveals white lupin candidate gene involved in anthracnose resistance. Theor Appl Genet 135(3):1011–1024. https://doi.org/10.1007/s00122-021-04014-7

Alkemade JA, Baroncelli R, Messmer MM, Hohmann P (2023) Attack of the clones: Population genetics reveals clonality of Colletotrichum lupini, the causal agent of lupin anthracnose. Mol Plant Pathol 24(6):616–627. https://doi.org/10.1111/mpp.13332

Alkemade JA, Patyi A, Arncken C, Lazzaro M (2025) Lupins: a remarkable protein crop battling anthracnose for a greener future. Plant Health Cases 2025:phcs20250008. https://doi.org/10.1079/planthealthcases.2025.0008

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402. https://doi.org/10.1093/nar/25.17.3389

Annicchiarico P, Osorio C, Nazzicari N, Ferrari B, Barzaghi S, Biazzi E, Tava A, Pecetti L, Notario T, Romani M, Crosta M (2025) Genetic variation and genome-enabled selection of white lupin for key seed quality traits. BMC Genomics 26(1):922. https://doi.org/10.1186/s12864-025-12048-0

Blighe K, Lun A (2023) PCAtools: PCAtools: everything principal components analysis. Doi:10.18129/B9.bioc.PCAtools (Version 2.14.0) [Computer software]. https://bioconductor.org/packages/PCAtools

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL (2012) Domain enhanced lookup time accelerated BLAST. Biol Direct 7:12. https://doi.org/10.1186/1745-6150-7-12

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421

Cerioli T, Hernandez CO, Angira B, McCouch SR, Robbins KR, Famoso AN (2022) Development and validation of an optimized marker set for genomic selection in southern U.S. rice breeding programs. Plant Genome 15(3):e20219. https://doi.org/10.1002/tpg2.20219

Chen Z-Q, Klingberg A, Hallingbäck HR, Wu HX (2023) Preselection of QTL markers enhances accuracy of genomic selection in Norway spruce. BMC Genomics 24(1):147. https://doi.org/10.1186/s12864-023-09250-3

CPVO (2026) *CPVO Variety Finder*. https://online.plantvarieties.eu/

Crossa J, Martini JWR, Vitale P, Pérez-Rodríguez P, Costa-Neto G, Fritsche-Neto R, Runcie D, Cuevas J, Toledo F, Li H, De Vita P, Gerard G, Dreisigacker S, Crespo-Herrera L, Saint Pierre C, Bentley A, Lillemo M, Ortiz R, Montesinos-López OA, Montesinos-López A (2025) Expanding genomic prediction in plant breeding: harnessing big data, machine learning, and advanced software. Trends Plant Sci 30(7):756–774. https://doi.org/10.1016/j.tplants.2024.12.009

de Mendiburu F (2023) agricolae: statistical procedures for agricultural research (Version 1.3-7) [Computer software]. https://cran.r-project.org/web/packages/agricolae/index.html

Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome. https://doi.org/10.3835/plantgenome2011.08.0024

Fu Y-B, Yang M-H, Zeng F, Biligetu B (2017) Searching for an accurate marker-based prediction of an individual quantitative trait in molecular plant breeding. Front Plant Sci. https://doi.org/10.3389/fpls.2017.01182

Gabur I, Chawla HS, Liu X, Kumar V, Faure S, von Tiedemann A, Jestin C, Dryzska E, Volkmann S, Breuer F, Delourme R, Snowdon R, Obermeier C (2018) Finding invisible quantitative trait loci with missing data. Plant Biotechnol J 16(12):2102–2112. https://doi.org/10.1111/pbi.12942

Hufnagel B, Marques A, Soriano A, Marquès L, Divol F, Doumas P, Sallet E, Mancinotti D, Carrere S, Marande W, Arribat S, Keller J, Huneau C, Blein T, Aimé D, Laguerre M, Taylor J, Schubert V, Nelson M, Péret B (2020) High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. Nat Commun 11(1):492. https://doi.org/10.1038/s41467-019-14197-9

Hufnagel B, Soriano A, Taylor J, Divol F, Kroc M, Sanders H, Yeheyis L, Nelson M, Péret B (2021) Pangenome of white lupin provides insights into the diversity of the species. Plant Biotechnol J 19(12):2532–2543. https://doi.org/10.1111/pbi.13678

Juliana P, Singh RP, Singh PK, Crossa J, Huerta-Espino J, Lan C, Bhavani S, Rutkoski JE, Poland JA, Bergstrom GC, Sorrells ME (2017) Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. Theor Appl Genet 130(7):1415–1430. https://doi.org/10.1007/s00122-017-2897-1

Kaler AS, Gillman JD, Beissinger T, Purcell LC (2020) Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Front Plant Sci 10:1794. https://doi.org/10.3389/fpls.2019.01794

Książkiewicz M, Nazzicari N, Yang H, Nelson MN, Renshaw D, Rychel S, Ferrari B, Carelli M, Tomaszewska M, Stawiński S, Naganowska B, Wolko B, Annicchiarico P (2017) A high-density consensus linkage map of white lupin highlights synteny with narrow-leafed lupin and provides markers tagging key agronomic traits. Sci Rep 7(1):15335. https://doi.org/10.1038/s41598-017-15625-w

Lenth RV, Banfai B, Bolker B, Buerkner P, Giné-Vázquez I, Herve M, Jung M, Love J, Miguez F, Piaskowski J, Riebl H, Singmann H (2025) emmeans: Estimated Marginal Means, aka Least-Squares Means (Version 1.10.7) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Luo C, Bashir NH, Li Z, Liu C, Shi Y, Chu H (2024) Plant micrornas regulate the defense response against pathogens. Front Microbiol. https://doi.org/10.3389/fmicb.2024.1434798

Meuwissen TH, Solberg TR, Shepherd R, Woolliams JA (2009) A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. Genet Sel Evol 41(1):2. https://doi.org/10.1186/1297-9686-41-2

Mierziak J, Wojtasik W (2024) Epigenetic weapons of plants against fungal pathogens. BMC Plant Biol 24(1):175. https://doi.org/10.1186/s12870-024-04829-8

Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA (2008) Database indexing for production MegaBLAST searches. Bioinf (Oxford England) 24(16):1757–1764. https://doi.org/10.1093/bioinformatics/btn322

Patyi A, Kamp M, Arncken C, Biazzi E, Książkiewicz M, Messmer MM, Schneider M, Tava A, Lazzaro M (2025) Identification of a new QTL associated to reduced quinolizidine alkaloid content in white lupin (*Lupinus albus*, L.) and development of ultra-low alkaloid recombinants by stacking with the pauper allele. BMC Plant Biol 25(1):945. https://doi.org/10.1186/s12870-025-06951-7

Phan HTT, Ellwood SR, Adhikari K, Nelson MN, Oliver RP (2007) The first genetic and comparative map of white lupin (*Lupinus albus* L.): identification of QTLs for anthracnose resistance and flowering time, and a locus for alkaloid content. DNA Res 14(2):59–70. https://doi.org/10.1093/dnares/dsm009

Puthiyedth N, Zeinalinesaz F, Hou D, Zhang Y, Lin W, Yan Y (2025) Leveraging LASSO-based methodologies for enhanced SNP analysis in plant genomes. Bioinform Adv 5(1):vbaf014. https://doi.org/10.1093/bioadv/vbaf014

R Core Team (2022) R: A language and environment for statistical computing. [Computer software]. R Found Stat Comput. https://www.R-project.org/

R Core Team (2024) R: A language and environment for statistical computing. [Computer software]. R Found Stat Comput. https://www.R-project.org/

Rychel-Bielska S, Nazzicari N, Plewiński P, Bielski W, Annicchiarico P, Książkiewicz M (2020) Development of PCR-based markers and whole-genome selection model for anthracnose resistance in white lupin (*Lupinus albus* L.). J Appl Genet 61(4):531–545. https://doi.org/10.1007/s13353-020-00585-1

Sallam AH, Endelman JB, Jannink J-L, Smith KP (2015) Assessing Genomic Selection Prediction Accuracy in a Dynamic Barley Breeding Population. Plant Genome 8(1):plantgenome2014.05.0020. https://doi.org/10.3835/plantgenome2014.05.0020

Schmidt P, Hartung J, Bennewitz J, Piepho H-P (2019) Heritability in plant breeding on a genotype-difference basis. Genetics 212(4):991–1008. https://doi.org/10.1534/genetics.119.302134

Schwertfirm G, Schneider M, Haase F, Riedel C, Lazzaro M, Ruge-Wehling B, Schweizer G (2024) Genome-wide association study revealed significant SNPs for anthracnose resistance, seed alkaloids and protein content in white lupin. Theor Appl Genet 137(7):155. https://doi.org/10.1007/s00122-024-04665-2

Sharma DK, Chatterjee M, Kaur G, Vavilala S (2022) 3—Deep learning applications for disease diagnosis. In D. Gupta, U. Kose, A. Khanna, & V. E. Balas (Eds.), Deep Learning for Medical Applications with Unique Data (pp. 31–51). Academic Press. https://doi.org/10.1016/B978-0-12-824145-5.00005-8

Skøt L, Nay MM, Grieder C, Frey LA, Pégard M, Öhlund L, Amdahl H, Radovic J, Jaluvka L, Palmé A, Ruttink T, Lloyd D, Howarth CJ, Kölliker R (2024) Including marker x environment interactions improves genomic prediction in red clover (Trifolium pratense L.). Front Plant Sci. https://doi.org/10.3389/fpls.2024.1407609

Talhinhas P, Baroncelli R, Le Floch G (2016) Anthracnose of lupins caused by colletotrichum lupini: a recent disease and a successful worldwide pathogen. J Plant Pathol 98(1). https://doi.org/10.4454/JPP.V98I1.040

Wang J, Zhang Z (2021) GAPIT version 3: boosting power and accuracy for genomic association and prediction. Genom Proteom Bioinform 19(4):629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wright S (1921) Correlation and causation. In J Agric Res 20(7):557. https://cir.nii.ac.jp/crid/1370567187556110595?lang=en

Yang H, Lin R, Renshaw D, Li C, Adhikari K, Thomas G, Buirchell B, Sweetingham M, Yan G (2010) Development of sequence-specific PCR markers associated with a polygenic controlled trait for marker-assisted selection using a modified selective genotyping strategy: A case study on anthracnose disease resistance in white lupin (*Lupinus albus* L). Mol Breed 25(2):239–249. https://doi.org/10.1007/s11032-009-9325-4

Yang X, Zhang L, Yang Y, Schmid M, Wang Y (2021) miRNA mediated regulation and interaction between plants and pathogens. Int J Mol Sci 22(6):2913. https://doi.org/10.3390/ijms22062913

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biology: J Comput Mol Cell Biology 7(1–2):203–214. https://doi.org/10.1089/10665270050081478

Zhang W, Boyle K, Brule-Babel A, Fedak G, Gao P, Djama Z, Polley B, Cuthbert R, Randhawa H, Graf R, Jiang F, Eudes F, Fobert P (2021) Evaluation of genomic prediction for Fusarium head blight resistance with a multi-parental population. Biology 10:756. https://doi.org/10.3390/biology10080756

Zhao T, Wang F, Qi J, Chen Q, Zhu L, Liu L, Yan L, Chen Y, Yang C, Qin J (2025) Genome-wide association analysis study and genomic prediction for resistance to soybean mosaic virus in soybean population. BMC Plant Biol 25:837. https://doi.org/10.1186/s12870-025-06775-5

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

András Patyi[1,2] · Michael Schneider[1] · Christine Arncken[1] · Michał Książkiewicz[2] · Monika M. Messmer[1] · Grit Schwertfirm[3] · Mariateresa Lazzaro[1]

✉ András Patyi
andras.patyi@fibl.org

[1] Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL), Ackerstrasse 113, Frick 5070, Switzerland

[2] Institute of Plant Genetics, Polish Academy of Sciences, Strzeszyńska 34, Poznań 60-479, Poland

[3] Institute for Crop Science and Plant Breeding, Bavarian State Research Center for Agriculture (LfL), Am Gereuth 2, Freising 85354, Germany

Supplementary information

Supplementary Tables in Publication nr. 2 (Patyi et al., 2026) include:

Supplementary Table S1: Samples included in the diversity panel, indicating the status of the accession, estimated marginal means phenotypic values, standard error and predicted phenotypic values as well as all genotyping available for the panel.

Supplementary Table S2: Samples included in the multi-parental panel, sAUDPC values from the field phenotyping, and predicted phenotypic values, letter coding of the pedigree, indicator of the generation, as well as all genotyping available for the panel.

Supplementary Table S3: Samples included in the multi-parental panel, estimated marginal means of phenotypic values from the climate chamber phenotyping, standard error and predicted phenotypic values as well as all genotyping available for the panel.

Supplementary Table S4: Samples included in the biparental panel, phenotypic values from a single late-season field phenotyping, and predicted phenotypic values, as well as all genotyping available for the panel.

Supplementary Table S5: Marker information of the panel of markers used for individual assessment indicating the chromosome and location of the SNP used for marker design, indicating reference and alternative allele at each loci, which primer type it is, the respective primer sequence, as well as marking which marker was transformed from a previous publication or was determined de novo for the study. In the latter case, also highlighting which model and genotyping was used for the marker definition and the explained variance.

Supplementary Table S6: Marker information of the panel of markers used for genomic prediction indicating the chromosome and location of the SNP used for marker design, indicating reference and alternative allele at each locus, marker effect, which primer type it is, the respective primer sequence, as well as marking which marker was transformed from a previous publication or was determined de novo for the study. In the latter case, also highlighting which model and genotyping was used for the marker definition.

Supplementary Table S7: Significance, explained variance, effect size and false discovery rate of each locus from the Analysis of Variance (ANOVA), including the number homo- and heterozygotes at each locus, and the respective anthracnose score means at each locus. "NA" refers to loci which were monomorphic in a given validation panel

All supplementary tables are available in the original published scientific article (https://link.springer.com/article/10.1007/s11032-026-01661-w#Sec20) and available for download at https://static-content.springer.com/esm/art%3A10.1007%2Fs11032-026-01661-w/MediaObjects/11032_2026_1661_MOESM1_ESM.xlsx

Declaration of the PhD student`s contribution in publications and the authors` declarations of their contribution

Publication nr. 2

Patyi, A., Schneider, M., Arncken, C., Książkiewicz, M., Messmer, M. M., Schwertfirm, G., Lazzaro, M. (2026). From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments. *Mol Breeding* 46, 37. https://doi.org/10.1007/s11032-026-01661-w

András Patyi, M.Sc.

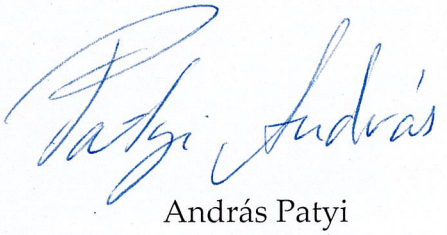
Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
5070 Frick
Switzerland

Institute of Plant Genetics, Polish Academy of Sciences
Strzeszyńska 34
60-479 Poznań
Poland

Declaration

I, András Patyi, declare that in the publication titled "From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments", my contributions are accurately declared in the original publication, based on the CRediT author statement system, as well as performed the duties of corresponding author:

"**A. P.:** Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing"



András Patyi

Frick, 2026.04.22.

Michael Schneider, Dr.

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
5070 Frick
Switzerland

Declaration

I, Michael Schneider, declare that in the publication titled "From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"**M. S.:** Conceptualization, Data curation, Formal Analysis, Investigation, Validation, Resources, Supervision, Writing – review & editing"



Michael Schneider

Frick, 2026.04.23.

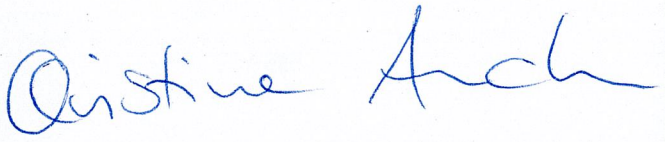
Christine Arncken, Dipl.-Ing. agr. ETH

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
Frick 5070
Switzerland

Declaration

I, Christine Arncken, declare that in the publication titled "From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"**C. A.:** Investigation, Project administration, Funding acquisition, field trial conceptualization, Resources, Writing – review & editing"



Christine Arncken

Frick, 2026.04.23.

dr hab. Michał Książkiewicz, prof. IGR PAN

Institute of Plant Genetics, Polish Academy of Sciences
Strzeszyńska 34
60-479 Poznań
Poland

Declaration

I, Michał Książkiewicz, declare that in the publication titled "From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"**M. K.:** Supervision, Writing – review & editing"

Michał Książkiewicz

Poznań, 2026.04.29.

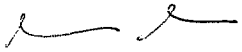
Monika M. Messmer, Dr.

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
5070 Frick
Switzerland

Declaration

I, Monika M. Messmer, declare that in the publication titled "From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments", my contributions are accurately declared in the original publication, based on the CRediT author statement system, as well as performed the duties of corresponding author:

"**M. M. M.:** Conceptualization, Funding acquisition, Project administration"



Monika M. Messmer

Frick, 2026.04.22.

Grit Schwertfirm, Dr.

Institute for Crop Science and Plant Breeding, Bavarian State Research Center for Agriculture (LfL)
Am Gereuth 2
85354 Freising
Germany

Declaration

I, Grit Schwertfirm, declare that in the publication titled "From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments", my contributions are accurately declared in the original publication, based on the CRediT author statement system:

"**G. S.:** Supervision, Writing – review & editing"

Grit Schwertfirm

Freising, 2026.04.23.

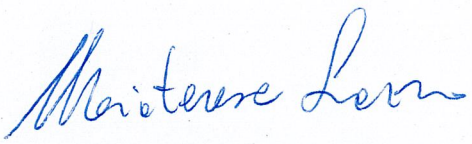
Mariateresa Lazzaro, Dr.

Department of Crop Sciences, Research Institute of Organic Agriculture (FiBL)
Ackerstrasse 113
Frick 5070
Switzerland

Declaration

I, Mariateresa Lazzaro, declare that in the publication titled "From research to application: Evaluation of literature-based and newly identified GWAS and GP-derived loci for anthracnose resistance in white lupin, across validation panels and environments", my contributions are accurately declared in the original publication:

"**M. L.:** Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing"



Mariateresa Lazzaro

Frick, 2026.04.22.

Statement of the thesis supervisors

We declare that this doctoral dissertation has been prepared under my supervision and we state that it meets the conditions to be presented in the procedure for granting the doctoral degree in the field of agricultural sciences in the discipline of agriculture and horticulture.

Date: 2026.05.22. Supervisor's signature: .....

Assoc. Prof. Michał Książkiewicz

scientific supervisor

Date: 2026.05.22. Supervisor's signature:…….

Dr. Mariateresa Lazzaro

auxiliary supervisor

Declaration of the author of the doctoral dissertation

Being aware of legal responsibility, I declare that the doctoral dissertation prepared by me does not include the content obtained in a manner inconsistent with applicable regulations. I also declare that the presented dissertation has not previously been the subject of procedures related to obtaining a scientific degree at a university or other authorized scientific institution. I further certify that this version of the dissertation is identical to the attached electronic version.

Date: 2026.05.22. Author's signature:……

András Patyi

PhD candidate