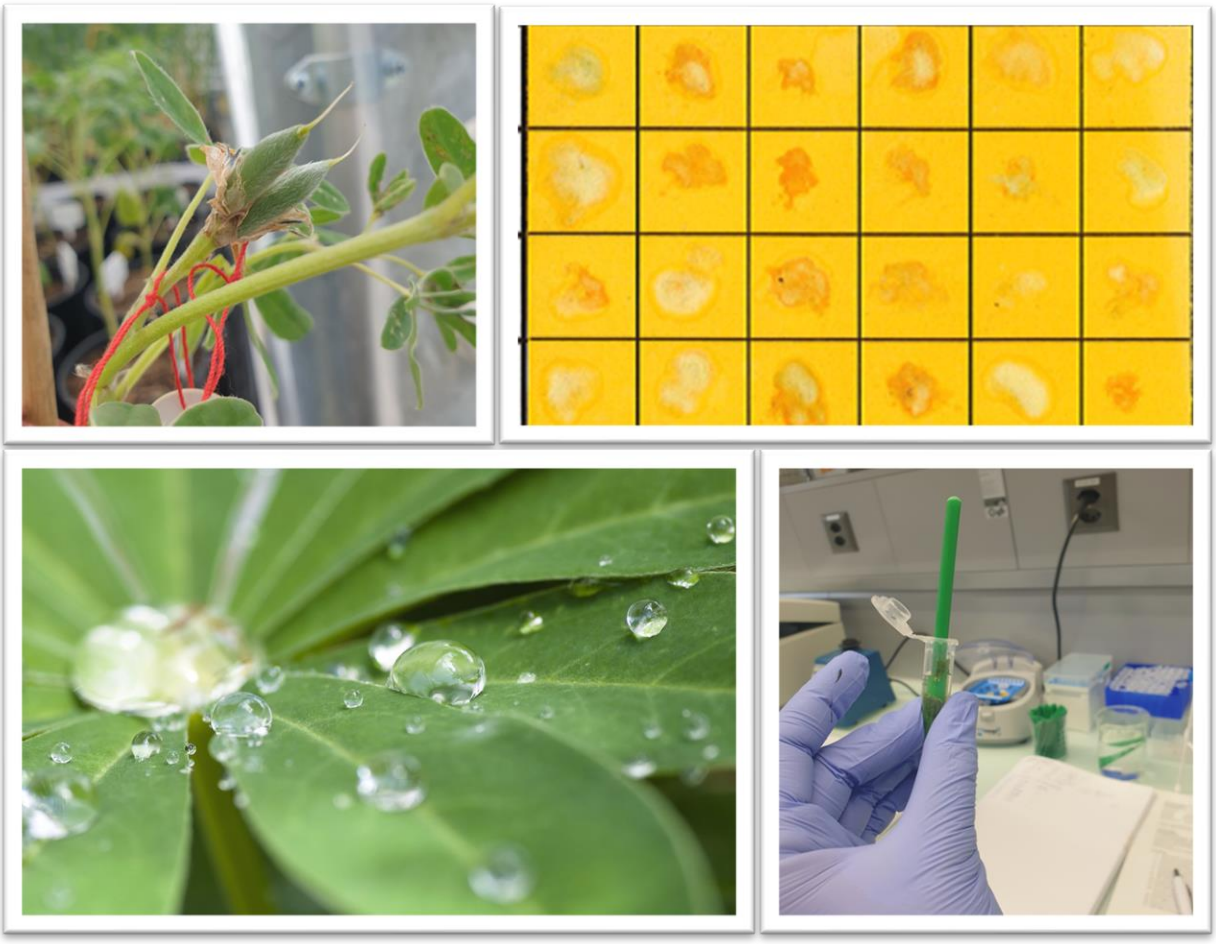
MSc Thesis

Breeding for low alkaloid content in white lupin (*Lupinus albus*)



Miriam Kamp, 23.01.2024

FiBL Switzerland

WAGENINGEN UNIVERSITY & RESEARCH

Breeding for low alkaloid content in white lupin (*Lupinus albus*)

Supervisors:

Dr. Chris Maliepaard (Wageningen University) & Dr. Mariateresa Lazzaro (FiBL Switzerland)

Miriam Inge Elisabeth Kamp | Student Number: 1050542

Master's Thesis | PBR80436 | 35 ECTS

Online Master's Plant Breeding (MPS-DL) | Wageningen University

Switzerland, 23.01.2024

Contents

General Introduction

White lupin (*Lupinus albus* L.) is a protein crop with high nutritional value and agronomic benefits like symbiotic nitrogen fixation, mobilisation of phosphate and adaptation to marginal acidic soils (Lucas et al., 2015; Talhinhas et al., 2016). The crop was domesticated in the Mediterranean basin and still is cultivated there as winter crop. When grown in Central Europe, it is used as a relatively cold tolerant summer crop (Lucas et al., 2015).

Despite the potential of white lupin, the cultivation in Central Europe is limited by several challenges, which can be, at least partly, targeted by breeding. Hence, important breeding targets in white lupins are:

- resistance to anthracnose disease (Talhinhas et al., 2016);
- early maturation as adaptation to Central European conditions (Lucas et al., 2015; Rychel et al., 2019);
- tolerance to calcareous soils to expand the potential cultivation area (Annicchiarico et al. 2023);
- minimized content of toxic and bitter quinolizidine alkaloids (QA) in the grains, to adhere with the recommended safety thresholds for use as feed and food (ANZFA, 2001);

A further important restraint for white lupin cultivation, which may be addressed by other means than breeding, is the novel and still underdeveloped market for lupin grains as feed. As an example, buyers and/or sellers in Germany often misjudge the feeding value of white lupin grains, which is comparable to that of soy, so that price offerings result to be grossly undervalued. Therefore, white lupin cultivation in Germany is relatively unprofitable and occurs mainly for on-farm feeding, not for trading. The limited supply in turn makes it difficult to develop a targeted demand for white lupins. However, as soon as market participants (especially producers) become aware of the feeding value of white lupins, prices may rise and with them the profitability of the crop (Tröster et al., 2023).

The research institute of organic agriculture FiBL (fibl.org) is dedicated to the advancement of organic farming in Europe. Since 2013, there are breeding and research activities for white lupin at FiBL Switzerland. The aim is to provide tools, knowledge and plant material to encourage white lupin breeding efforts in Europe. The research is focussed on anthracnose resistance, early maturity and low alkaloid content. FiBL collaborates with gzpk (gzpk.ch), a non-profit organic breeding association in the scope of a current research project (LUPINNO Suisse, 2021-2024, funded by the Swiss Federal Office for Agriculture). Current breeding activities at FiBL are targeted at pyramiding different alleles for a low alkaloid content into germplasm with increased anthracnose resistance, in order to develop cultivars with high stability for yield and minimized alkaloid contents. This thesis aims specifically to improve understanding of alkaloid genetics in white lupin, to facilitate breeding for this aspect. This thesis was conducted within the scope of the EU research project DIVINFOOD.

In this work, the genetic architecture of alkaloid content in white lupin was investigated from many different angles and by different means. This comprised data collection from segregating populations in the field, performing crosses, conducting different experiments with the F1 seeds, isolation of DNA with different methods for different purposes, performing PCR based genotyping, application of different statistical approaches to analyse and interpret phenotyping and genotyping data. To present and contextualise the results, this thesis is divided into two chapters, which were written in the style of individual journal articles. In the first chapter the investigation of maternal and cytoplasmic effects on QA levels white lupin is presented. The second chapter is dedicated to the validation of molecular markers for QA levels and the mapping of a novel QTL in white lupin. The thesis is rounded off by a general conclusion merging the key findings from both chapters.

The aims of this thesis, and the assignment to chapters I and II are as follows:

a) Exploration whether quinolizidine alkaloids are transported to or synthesized in the grains (chapter I);
b) Quantification of cytoplasmic effects on alkaloids contents (chapter I);
c) Validation of candidate SNP markers for low alkaloid content (chapter II);
d) Analysis of genotypic and phenotypic segregation patterns to quantify individual and combined effects of low alkaloid genes (chapter II)
e) Bulked segregant analysis to map a new quantitative trait locus (QTL) associated with alkaloid content (chapter II)

Literature

Annicchiarico, P., de Buck, A. J., Vlachostergios, D. N., Heupink, D., Koskosidis, A., Nazzicari, N., & Crosta, M. (2023). White Lupin Adaptation to Moderately Calcareous Soils: Phenotypic Variation and Genome-Enabled Prediction. *Plants*, *12*(5), Article 5. https://doi.org/10.3390/plants12051139

ANZFA. (2001). *Lupin Alkaloids in Food A Toxicological Review and Risk Assessment*. Australia New Zealand Food Authority. https://www.foodstandards.gov.au/publications/documents/TR3.pdf

Lucas, M. M., Stoddard, F., Annicchiarico, P., Frias, J., Martinez-Villaluenga, C., Sussmann, D., Duranti, M., Seger, A., Zander, P., & Pueyo, J. (2015). The future of lupin as a protein crop in Europe. *Frontiers in Plant Science*, *6*. https://www.frontiersin.org/articles/10.3389/fpls.2015.00705

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

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), 5–14. https://doi.org/10.4454/JPP.V98I1.040

Tröster, M. F., Heinz, M., & Durst, L. (2023). Determining the Value of Novel Feedstuffs in Imperfect Markets, Taking Lupinus albus as an Example. *Agriculture*, *13*(4), Article 4. https://doi.org/10.3390/agriculture13040867

Chapter I: Alkaloids in white lupins are transported to the seeds and their levels are partly inherited through the cytoplasm

Abstract

White lupin grains contain quinolizidine alkaloids (QA), which are bitter and toxic compounds. Both historically and in the future, it continues to be an important breeding target to minimise the QA contents in white lupin. In this study, the maternal effect on QA accumulation in the seeds was investigated by comparing QA contents in selfed and F1 seeds formed on the same maternal genotype. Furthermore, we explored the potential of cytoplasmic effects on QA contents for breeding by measuring and comparing QA levels in F1 plants from reciprocal crosses. Our data confirms, that QA content in white lupin seeds is determined by the genotype of the mother plant. The results from the reciprocal F1 crosses suggest that cytoplasmic inheritance has an effect on the QA content in white lupin. Further research is required to elucidate whether the cytoplasmic effects are transmitted paternally or maternally, because in this regard our results were inconclusive while previous studies showed conflicting results.

Introduction

White lupins (*Lupinus albus* L.) are grain legumes suited for local protein production in Central Europe, especially for low-input and organic agriculture, due to the high protein content of their grains and their abilities of mobilizing phosphorus, fixing nitrogen and withstanding cold and drought (Lucas et al., 2015). Still, widespread cultivation is hampered on the one hand by the destructive ascomycete *Colletotrichum lupini* causing anthracnose disease (Talhinhas et al., 2016), and on the other hand by the seeds' content in quinolizidine alkaloids (QAs). QAs have a bitter taste and are toxic at higher concentrations. Therefore, genotypes with high QA contents (hQA) are commonly referred to as "bitter" and genotypes with low QA contents (lQA) are in distinction referred to as "sweet" (Mancinotti et al., 2023). The recommended safety thresholds for QA levels in lupin grains are 500 mg kg^{-1} for feed and 200 mg kg^{-1} for human consumption (ANZFA, 2001). Since QAs are water-soluble, it is traditional practice to debitter lupin grains with soaking procedures to use the softened seeds in brine as snack (Erbas 2010). However, this practise is economically and environmentally challenging for large scale processing of lupin grains and flour.

Seeds of bitter white lupin cultivars contain about 2'000-4'000 mg kg^{-1} and seeds of extremely bitter wild material up to 127'000 mg kg^{-1} (Kroc et al., 2017). There are different genes that cause reduced accumulation, but not fully eradicate QAs in white lupin seeds, with the lQA phenotype always being recessive. Most market cultivars possess at least one lQA gene and thus are considered "sweet". The most deployed lQA gene is at the *pauper* locus, and was characterised to reduce the seed alkaloid content to 200-500 mg kg^{-1} (Harrison & Williams, 1982). The causal point mutation for *pauper* was recently identified and its effect shown to derogate the function of an acetyltransferase, which catalyses an early step of QA biosynthesis (Mancinotti et al., 2023). Besides *pauper* there are other lQA genes in white lupin, described to reduce the alkaloid content to ≤1000 mg kg^{-1} (*mitis*, *reductus*, *exiguus*, *nutricius*), as well as further genes (*suavis*, *minutus*) of which the effects on QA levels have not yet been quantified (Harrison & Williams, 1982).

The QA levels in grains of "sweet" cultivars have been observed to fluctuate and often exceed the advised thresholds, *e.g.* in response to abiotic stress (Annicchiarico et al., 2014). This represents a substantial economic risk for farmers and producers. Therefore, it is desirable to minimise the QA levels in white lupin by breeding, so that the safety threshold values are complied with, regardless of stress-caused fluctuations.

One cause for the observed fluctuations of QA contents in white lupin grains could be outcrossing by insect pollination, resulting in bitter F1 seeds (xenia). This would be relevant only if QA accumulation in seeds were influenced by the genotypes of endosperm or embryo. In the related species *Lupinus angustifolius* L. it was proven that QAs are synthesised in maternal tissue and deposited into the developing seeds, regardless of the F1 (embryo) genotype of the seed (Otterbach et al., 2019). A recent study suggests accordingly for white

lupin, that outcrossing does not affect the QA content in the grains, based a study where the cultivar Feodora (*pauper*) which is fairly sweet (226 mg kg^{-1}) and the bitter cultivar Mihai (11.981 mg kg^{-1}) where grown in isolation tunnels with different treatments encouraging (by inclusion of bumblebees) and preventing (by exclusion of pollinators) cross pollination (Cuijpers et al., 2023; Cuijpers & Heupink 2022). The rate of outcrossing was however not measured. In an older experiment, no effect of the father plant on the QA level in lupin seeds could be detected, however the quantification method at that time was not as precise as today's methods and the QA quantity of seed from "sweet" cultivars could not be accurately calculated, so that the statistical analyses were conducted by comparison of the total area of G.L.C peaks (Harrison 1980).

One way to breed for lower QA contents and reduce the likelihood of exceeded QA thresholds could be to combine two or more lQA genes. Since lQA genes are recessive and complementary, F2 populations segregating for two lQA genes segregate phenotypically with 7:9 into "sweet" and bitter plants (Harrison et al., 1982; Porsche 1964). In this population, only 1:15 plants are homozygous for both lQA genes, which is challenging to select for, especially in breeding programs with complex crosses. Another strategy to minimize QA contents in white lupin could be to combine a lQA gene (e.g. *pauper*) with a maternally transmitted cytoplasmatic source for sweetness. The advantage of exploiting maternal cytoplasmic effects in combination with e.g. *pauper* over combining multiple nuclear lQA genes, is the maternal transmission, resulting in less complicated segregation patterns and easier traceability in a breeding program. Regarding the existence of cytoplasmic effects on QA contents, contrasting observations were made based on alkaloid phenotypes of F1 plants from reciprocal crosses. For her dissertation, Harrison (1980) observed in three out of 14 reciprocal cross-combinations significant differences in QA content, of such magnitude, that there was no overlap of the data ranges of the reciprocal F1s. Harrison attributed these effects to a paternal effect of the cultivars Gyulatania and Shinfield which were present in those three cross combinations. Conversely, the presented data could also be interpreted as maternal effect of the cultivars Kievskij mutant and Kievskij Skorospely. In a different experiment in her dissertation, Harrison detected significant differences in reciprocals of a full diallel cross of the cultivars Nährquell, Neuland and Ultra, suggestive of a maternal transmission of the effect (Harrison 1980). The magnitude of the maternal cytoplasmic effect accounted for 320-940 mg kg^{-1} difference between reciprocals. For this experiment it is not clear, whether the replicates were biological or pseudo replicates, and which statistical test was used. In a later published experiment involving reciprocal crosses, no cytoplasmic effects were reported whatsoever (Harrison & Williams, 1982).

Cytoplasmic effects have been used as means for breeding, *e.g*. variation of mitochondrial DNA (mtDNA) was shown to elevate the level of unsaturated fatty acids and increase drought tolerance in sunflower. Furthermore, photosynthetic efficiency and yield in rice and flower color in petunia had been shown to be influenced by mtDNA (Yoosefzadeh-Najafabadi et al., 2023). Chloroplastic DNA (cpDNA) from *lupinus luteus* has been used for for phylogenetic studies of fabaceae (Martin et al. 2014) and variation in cpDNA can be used in breeding to check for hybridisation (Park et al., 2021; Yoosefzadeh-Najafabadi et al., 2023).

This study aims to validate maternal effects on QA accumulation in white lupin seeds, by performing crosses and with modern analytical quantification of QAs. We hypothesise that the QA content of seeds solely depends on the maternal genotype. We used the bitter genotype Weibit, the cultivar Butan which carries the lQA gene *exiguus* and the inbred line exMurringo (*pauper*) which was derived from an off-type single plant in the cultivar Murringo and is extremely sweet. To confirm that QAs are transported from maternal tissue to the seeds in *L. albus,* the QA contents in F1 seeds and selfed seeds from the same parental plants were measured by gas chromatography (GC) and compared.

Moreover, this study aims to infer, whether genetic variation of the cytoplasmic DNA has an effect on the QA levels in white lupin. For this, the QA contents were measured and compared for F1 plants from reciprocal crossings, assuming no phenotypic difference between reciprocal F1 plants with the same parental genotypes.

Material and Methods

Crosses

Three accessions contrasting for their QA contents and genotypes, were crossed in full diallel: The hQA genotype Weibit, the lQA genotype Butan with the gene *exiguus* and the lQA line exMurringo carrying the gene *pauper*. Due to the complementary and recessive sweetness genes present in exMurringo and Butan, all cross combinations should result in bitter F1 plants. The phenotypes of all crossing parents were tested by Dragendorff test and some bitter Butan plants were excluded. The Dragendorff test was conducted by squeezing sap from the petioles on Dragendorff paper (Plarre 1985). The Dragendorff paper had been prepared by infusing filter paper with Dragendorff solution (iodine/potassium iodide solution) and subsequent drying at 40°C. A visible color change of Dragendorff solution to a reddish brown indicates QA levels ≥ 500 mg kg^{-1} (Li et al., 2011), whereas at low QA levels in the sap, the bright yellow colour of the Dragendorff paper fades to a white stain.

The cultivars were sown sequentially in several batches to allow crosses of genotypes with different flowering time. The batches consisted of ten plants per genotype and were sown with ten days offset and the date of the first sowing was the 30/1/2023. In the first batch, only the genotypes ExMurringo and Weibit were sown (since Butan is an early flowering type), while all other batches comprised all genotypes. The plants were sown in 2L pots with substrate composed of 4/5 universal potting substrate mixed with of 1/5 vermiculite. The seedlings were inoculated by mixing a few grams of peat-based bradyrhizobia inoculant (LEGUMEFiX®) with the first watering. The plants were grown in a greenhouse with 12h light per day and with daily watering by automatic flooding tables. The greenhouse can be heated but not cooled, and the temperature was set to 15°C at minimum. Due to a warm and sunny spring, the temperatures in the greenhouse reached 30°C in march. These conditions likely favoured infestations with thrips, aphids and spider mites, which were treated against by showering the plants, spraying with spinosad and azadirachtin (NeemAzal®), and release of predatory mites (with due time interval from insecticide use). The crosses were conducted by emasculating flowers to prevent self-pollination, and pollination with the crossing partner with one to two days delay. For plants with developing F1 pods, all except for two selfed pods per plant were nipped. The F1- and selfed seeds were harvested when the pods where fully ripe. The thrips infestation was very damaging and reduced the number of seeds that could be harvested from the plants, especially from the genotype ExMurringo.

Sampling

Samples of the F1-seeds and corresponding selfed seeds from the same maternal genotypes were selected, with three biological replicates per cross combination. Every sample consisted of two seeds, and when possible, the F1 seeds and the paired sample of selfed seeds were selected to have developed on the same plant individual(s). When no corresponding selfed seeds from the same mother plant were available, selfed seeds from other plants of the same genotype were used (Suppl. Table I - 1).

For the second experiment, F1 seeds from the reciprocal crosses of Butan, Weibit and ExMurringo were sown in a randomised complete block design (RCBD) with 6 blocks on the 20/09/23. Due to the thrips infestation, there were no seeds for ExMurringo × Weibit, and hence no F1 seeds were sown for the reciprocals of this cross. The same pots and substrate were used as described for the crossings, however with slow-releasing fertiliser added to the substrate. The experiment was set up in a climate chamber with 16h daylength at 23-25°C. The light was first set to blue containing light (58W fluorescent tube light colour 840 "cool daylight") to induce vegetative biomass production and 30 days after sowing changed to red containing light (sodium vapor lamps, 400 W) to induce flowering.

To include only true F1s and exclude plants that were selfed unintendedly, DNA samples were collected from all F1 plants and the parental genotypes with Whatman FTA cards and sent to LfL (Bavarian State Research Center for Agriculture) for a genetic success control. At LfL, the

DNA was eluted from the FTA cards and genotyping was conducted with a set of 96 SNP markers spread over the genome in a Standard BioTools™ genotyping array.

Pest control involved treatments of the substrate with B.t.i (Solbac®) against *Sciaridae* and Spinosad preventative against thrips. The plants were watered as needed, but with always the same quantity given to all plants at once. The samples were collected from flowering plants at the exact same phenological state. For this, the plants were checked daily and the samples harvested, when the uppermost flower of their main inflorescence had opened and the lowest 2 florets started wilting (BBCH 65). The sample harvest spanned a period from the 26/10 to 12/11/23. The samples were taken by cutting off the main inflorescence and two young, yet fully expanded leaves. Since the QA content in plant leaves fluctuate during the day and peak in the afternoon (Wink & Witte, 1984), the sample collection was conducted always at 3:30 PM.

Alkaloid analytical quantification

The seeds and the dried plant tissue samples were milled with a Quiagen TissueLyser at max speed (30 Hz) for 30 seconds. The alkaloid content of the seeds and plant material was analysed with GC-MS at CREA-ZA (Research Centre for Animal Production and Aquaculture, Italy). For QA extraction, GC-MS and quantification, a modified version of the protocol by Boschin et al. (2008) was used, which was adapted to the specifics of CREA-ZA laboratory. With this method, the following 10 compounds can be detected to a level as low as 0.3 mg kg^{-1} : Ammodendrine, tetrahydro hombifoline, angustifoline, a-isolupanine, lupanine, 11,12-seco-12,13-didehydro multiflorine, multiflorine, 13a-hydroxy lupanine, 13a-angeloyl oxylupanine and 13a-tigloyl oxylupanine.

Data analyses

All data analyses were conducted using the R environment for statistical computing, version 4.3.1. For each maternal genotype, pairwise t-tests were used to test, whether the differences in QA content of selfed seeds and F1 seeds significantly differ from zero. The paired treatments per genotype (and therefore per t-test) consisted of three biological replicates with two technical replicates respectively.

The QA contents of the F1 plants were analysed for differences between the reciprocal versions of the same cross (with the same nuclear genotypes). There were 24 biological samples and two technical replicates. The variances were heterogeneous over the two different nuclear genotypes. A linear model was fitted with the function gls() from the package nlme was used (v3.1-152; Pinheiro et al., 2021), by correcting for heterogeneity of variance (with the parameter *weights = varIdent(form =~ 1| nuclear_genotype)*. There was no significant block effect (RCBD) and hence the blocks were not used as covariate in the linear model (Suppl. Figure I - 1). The interaction between nuclear genotype and motherplant could not be tested, because the gls was rank deficient (probably because there were not enough treatment combinations to compute interaction terms). The resulting model was:

total_ppm ~ nucleargenotype+motherplant, weights = varIdent(form =~ 1|nucleargenotype).

Subsequently, pairwise comparisons were conducted with tukey tests using *glht()* from the package multcomp (v1.4-16; Hothorn et al., 2008) and the residuals were checked graphically (Suppl. Figure I - 2).

Results

Xenia: effect of pollen genotype on alkaloid content of F1 seeds

The total seed QA content in the 36 analysed samples ranged from 13.73 to 4412.37 mg kg^{-1}.

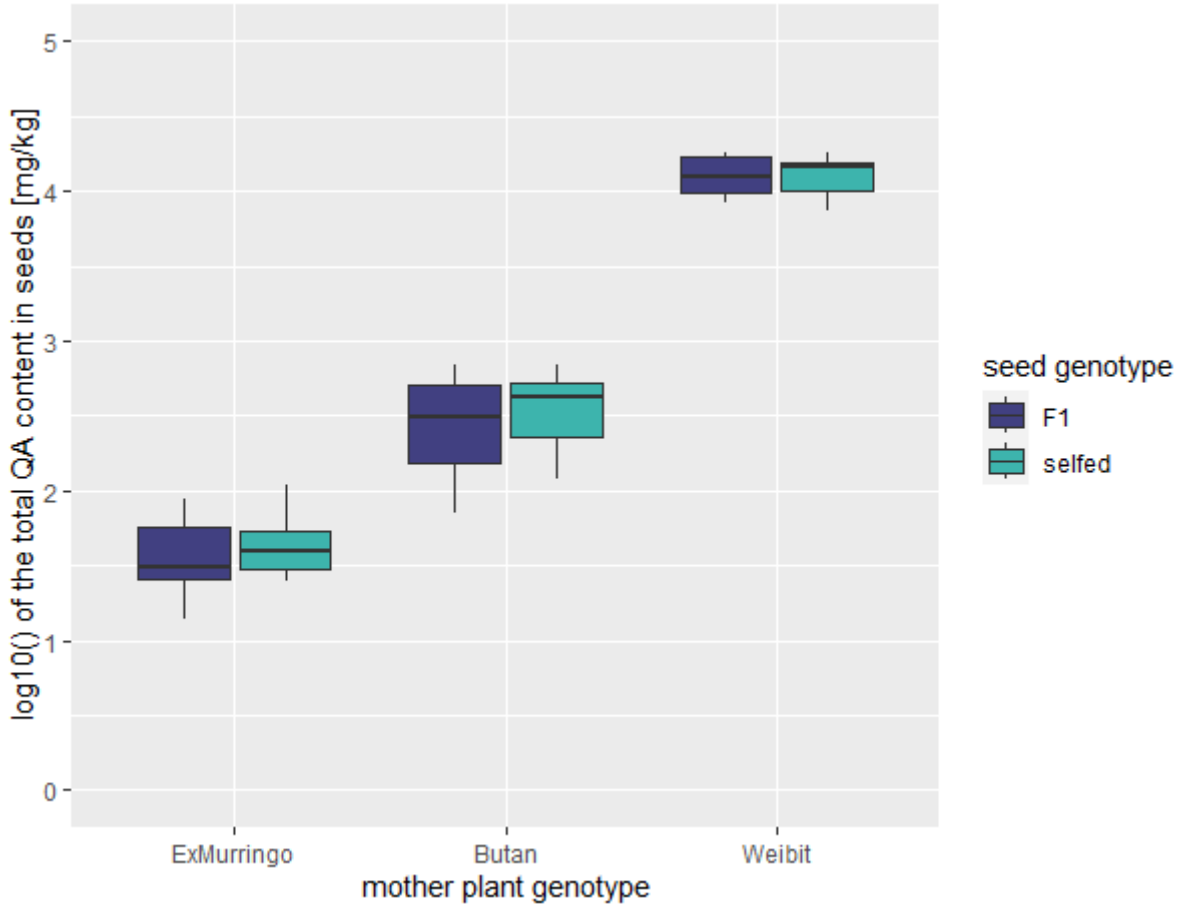


Figure 1: Decadic logarithm of QA contents (mg kg^{-1}) in white lupin seeds by mother plant and seed embryo genotypes, which were either cross-pollinated (F1-seeds) or selfed. The decadic logarithm was used for visualizing the comparison of QA contents in selfed and F1 seeds for each maternal genotype, despite a more than 300-fold difference in total QA content.

The genotypic means for the total QA contents (selfed seeds only) was 48 mg kg^{-1} for ExMurringo, 388 mg kg^{-1} for Butan and 13023 mg kg^{-1} for the bitter cultivar Weibit. There was no significant difference in QA content between selfed and F1 seeds for any of the three mother plant genotypes (Table 1).

Table 1: Paired t-test statistics for comparison of QA contents in selfed and F1 seeds by mother-plant genotype

Mother plant genotype	Compared seed genotypes	mean difference for QA content (mg kg^{-1})	df	t	p-value
ExMurringo	Selfed vs. F1	7.60	11	0.62	0.55
Butan	Selfed vs. F1	30.45	11	0.75	0.47
Weibit	Selfed vs. F1	191.95	11	0.10	0.92

Female and male parental effect on alkaloid content in F1 plants of reciprocal crossings

All F1 plants were heterozygous at loci contrasting for the parentals and therefore it can be concluded, that the crosses were successful and no unintended selfing occurred (Suppl. Figure I - 3). The total QA content in the F1 plants ranges from 577.4 mg kg^{-1} to 10295.3 mg kg^{-1}.

The means for the two different nuclear genotypes are 1500 mg kg^{-1} for the reciprocal crosses of Butan and ExMurringo (BxE_ExB) and 7360 mg kg^{-1} for the reciprocal crosses of Butan and

Weibit (BxW_WxB). These values exceed the threshold levels of 200 and 500 mg kg^{-1} for food and feed respectively (ANZFA, 2001) and hence all plants can be considered "bitter". The results for the statistics of the mother plant effects are shown in Table 2.

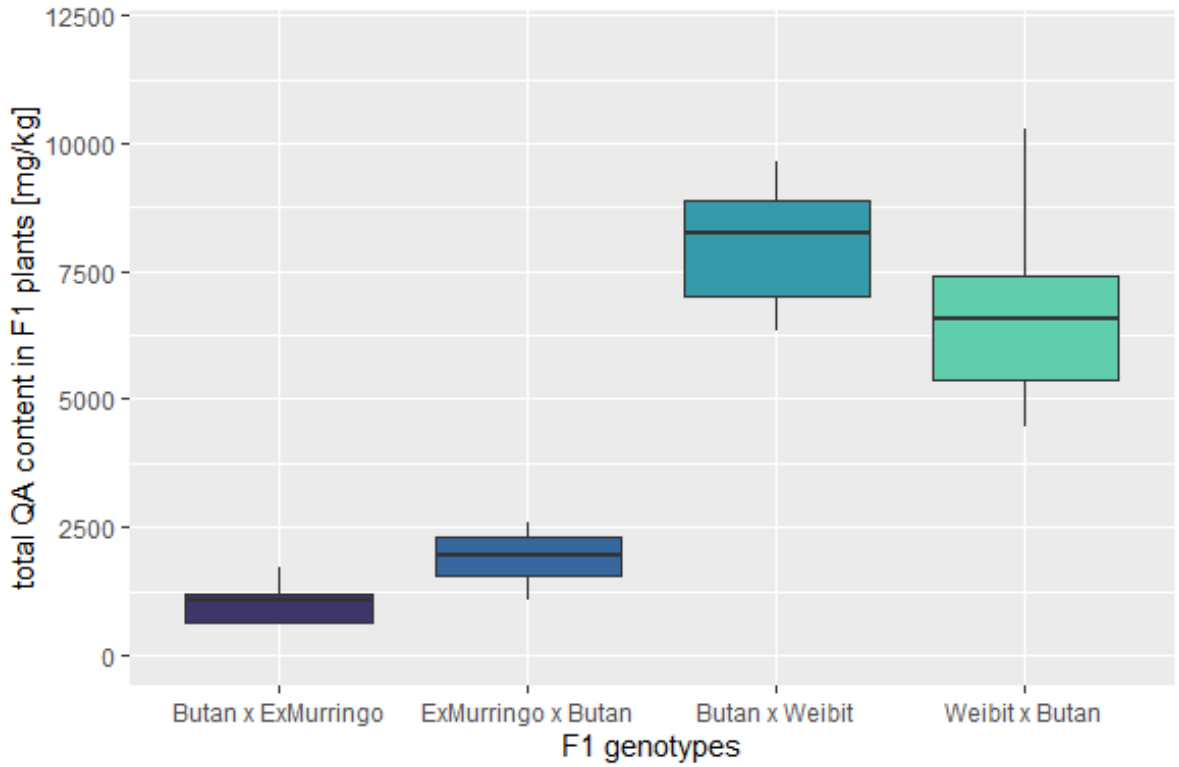


Figure 2: Total QA contents (mg kg^{-1}) in the plant tissue of flowering F1 plants by reciprocal crosses.

Table 2: Pairwise comparisons (tukey) for the effect of the cytoplasm and the nuclear genotypes on QA content in F1 plants derived from reciprocal crosses. ExB_BxE indicates the nuclear genotype, which is the same for both reciprocal crosses of ExMurringo and Butan. BxW_WxB indicates the nuclear genotype for the F1 plants for the crosses of Butan and Weibit.

pairwise comparisons of cytoplasmic effects	**estimated effect on total QA content (mg kg^{-1})**	**standard error**	**z value**	**p-value**
ExMurringo - Butan	882.4	185.1	4.767	<0.001
Butan - Weibit	1228.3	590.1	2.083	0.0818
pairwise comparison of nuclear genotypes	**estimated effect on total QA content (mg kg^{-1})**	**standard error**	**z value**	**p-value**
ExB_BxE vs. BxW_WxB	6951.2	437.7	15.88	<0.0001

Discussion

The results of this study confirm that the QA content in white lupin grains is determined by the genotype of the mother plant. It is likely that, as described for narrow leafed lupin (Otterbach et al., 2019), the QAs are produced in green vegetative tissue and deposited into the developing seeds. This finding is consistent with results reported by Harrison (1980) in her PhD dissertation and Cuijpers et al. (2023). For breeding, these results imply that by measuring the QA content in seeds, the previous generation is phenotyped. The implication for white lupin cultivation is that outcrossing on the field will not cause increased QA levels in the harvested produce and that different cultivars of white lupin may be grown on neighbouring parcels as long as the grain is not used for farm-saved seed. Nevertheless, this also highlights the importance of isolation of white lupins during seed multiplication, since,

given the outcrossing rate 8-10% (Green et al., 1980 as cited by Hufnagel et al. 2021), seed batches with low QA contents may result in bitter plants and bitter grain yields in the next generation.

The fact that all F1 plants were bitter is consistent literature describing all lQA alleles to be recessive (Harrison and Williams, 1982; Swiecicki et al. 2019). Furthermore, the high QA contents in the F1 plants derived from crossing the "sweet" cultivars ExMurringo and Butan is consistent with the description of Butan to carry *exiguus* (Rychel et al., 2019) and of ExMurringo to carry *pauper*. The two sweetness genes are complementary and hence heterozygous in the F1, which results in a bitter phenotype due to the recessive nature of "sweetness".

The QA level in the reciprocal crosses for Butan and ExMurringo is much lower than in the other reciprocal crosses, although no major lQA gene is homozygous in either F1. Differences in QA levels that cannot be explained by absence or presence of major lQA genes have been documented before. Porsche (1964) described 2 to 3-fold differences in breeding progenies ("Stämme") that were all carrying the sweetness gene *pauper* and attributed these differences to the genetic backgrounds or involvement of modifier genes. Another explanation may be incomplete recessiveness of sweetness genes, which had been observed in segregating populations by Harrison and Williams (1982). Regarding the QA contents in F1 plants from reciprocal crosses, there is a significant difference for the crosses of Butan and ExMurringo with lower QAs with Butan as female parental, and no significant differences between the crosses of Butan and Weibit but with a conflicting tendency with higher QAs with Butan as female parental.

There are two different ways how the cytoplasm may explain the observed data: maternal and paternal cytoplasmic effects. Firstly, maternal effects may explain the significant difference between the reciprocal crosses of Butan and ExMurringo. This would align with the observation of cytoplasmic effects in white lupin in chapter 4 of the dissertation of Harrison (1980). The magnitude of the observed significant difference between reciprocals of the cross of Butan and ExMurringo was powerful, with a reduction of estimated 882.4 mg kg^{-1}, resulting in QA contents of 1036 mg kg^{-1} for the F1 with Butan as maternal parent. This difference is substantial, however occurs at very high QA levels, considering the recommended thresholds of 500 and 200 mg kg^{-1}. It is impossible to extrapolate this magnitude to plants with very low QA contents and hence determine the use of the effect for breeding white lupins with lower QA contents.

However, with maternal cytoplasmic effects, it is puzzling that ExMurringo as mother plant resulted in higher QA levels than with Butan, while the cultivar ExMurringo has lower QA levels. Still, this needn't contradict the presence of a cytoplasm effect, since the observed trend may be by chance or this might hint at presence of an interaction of cytoplasm and nuclear genotype. If an interaction like this were present, it would complicate the applicability of cytoplasmatic QA reduction in breeding. Interaction between cytoplasm and nuclear genes is a well-known phenomenon, for example in cytoplasmatic male sterility systems in sunflower, maize, cabbage and other crops for hybrid breeding and hybrid seed production (Eckardt 2006, Wang et al. 2011).

Conversely, the observed data and trends may be explained by the opposite phenomenon: paternal effects. In chapter 3 of her dissertation, Harrison (1980) observed very strong and evident paternal effects on QA contents of F2 seeds (which reflect the phenotype of the F1). An effect of the pollinator genotype to explain the observed differences between the reciprocal crosses would explain the observed trends with better coherence with the QA phenotypes of the three cultivars. Unfortunately, there were no F1 plants for the crosses ExMurringo × Weibit and Weibit × Exmurringo, due to a thrips infestation in the greenhouse during the crosses. Hence, the dataset is not appropriate to capture pollinator effects, as only Butan is present twice as pollinator in the F1 plants. The inheritance of organelles in plants can occur both maternally (more common) and paternally. In the model legume *Medicago truncatula*, plastids are inherited biparentally (Matsushima et al. 2008). In *nicotiniana tabacum*, plastids

are transmitted maternally. However, mild chilling stress during pollen formation strongly increased the paternal plastid transmission in *nicotiana* (Chung et al., 2023). Environmental effects could hence account for the inconsistencies in earlier studies regarding cytoplasmic effects (Harrison 1980, Harrison and Williams 1982).

Be it due to maternal or paternal effects, the data suggests that the direction of crosses seem to be relevant in white lupin breeding. As the exact mechanism of parental effects is uncertain, reciprocal crosses should be considered to improve the chances of breeding progress regarding QA content. The F2 seeds or F3 generation from the investigated plants may furthermore be used to validate the observed effects in the sweet and bitter segregating plants, to improve the understanding of whether the observed effects are measurable at low QA levels and stable over the generations, as cytoplasmic effects should be. Furthermore, in future studies, reciprocal versions of crosses with genotypes carrying the same lQA gene, *e.g. pauper* × *pauper* or *exiguus* × *exiguus* should be included to measure the magnitude of the effect at low QA levels.

Conclusion

We can conclude that the QA contents in white lupin grains are determined by the genotype of the mother plant. Furthermore, the direction of crosses seems to be relevant, although the role of the parental plants and organelles in determining differences between reciprocals remains to be unravelled by future research. The cytoplasmic effect is strong at high QA levels, and should this magnitude be sustained at lower QA levels, it might represent a powerful means to reduce QA contents in white lupin breeding for cultivars with minimised QA contents.

Literature

Annicchiarico, P., Manunza, P., Arnoldi, A., & Boschin, G. (2014). Quality of *Lupinus albus* L. (White Lupin) Seed: Extent of Genotypic and Environmental Effects. *Journal of Agricultural and Food Chemistry*, *62*(28), 6539–6545. https://doi.org/10.1021/jf405615k

ANZFA. (2001). *Lupin Alkaloids in Food A Toxicological Review and Risk Assessment*. Australia New Zealand Food Authority. https://www.foodstandards.gov.au/publications/documents/TR3.pdf

Boschin, G., Annicchiarico, P., Resta, D., D'Agostina, A., & Arnoldi, A. (2008). Quinolizidine Alkaloids in Seeds of Lupin Genotypes of Different Origins. *Journal of Agricultural and Food Chemistry*, *56*(10), 3657–3663. https://doi.org/10.1021/jf7037218

Chung, K. P., Gonzalez-Duran, E., Ruf, S., Endries, P., & Bock, R. (2023). Control of plastid inheritance by environmental and genetic factors. *Nature Plants*, *9*(1), Article 1. https://doi.org/10.1038/s41477-022-01323-7

Cuijpers, W., & Heupink, D. (2022). *Bestuiving en oogstzekerheid in eiwitgewassen*. Louis Bolk Instituut.

Cuijpers, W. J. M., Heupink, D., Mulder, P. P. J., & Prins, U. (2023). Effect of cross-pollination of sweet and bitter lupins on alkaloids, effect of pollinators on yield and pollinator diversity in lupins. *„Breeding, Cultivation and Use of Lupins for a Sustainable Agriculture – Recent Developments"*, 17.

Czepiel, K., Krajewski, P., Wilczura, P., Bielecka, P., Święcicki, W., & Kroc, M. (2021). Expression Profiles of Alkaloid-Related Genes across the Organs of Narrow-Leafed Lupin (Lupinus angustifolius L.) and in Response to Anthracnose Infection. *International Journal of Molecular Sciences*, *22*(5), Article 5. https://doi.org/10.3390/ijms22052676

Eckardt, N. A. (2006). Cytoplasmic Male Sterility and Fertility Restoration. *The Plant Cell*, *18*(3), 515–517. https://doi.org/10.1105/tpc.106.041830

Erbas, M. (2010). The Effects of Different Debittering Methods on the Production of Lupin Bean Snack from Bitter Lupinus Albus L. Seeds. *Journal of Food Quality*, *33*(6), 742–757. https://doi.org/10.1111/j.1745-4557.2010.00347.x

Fischbeck, Plarre, & Schuster. (1985). *Lehrbuch der Züchtung landwirtschaftlicher Kulturpflanzen* (2nd ed.).

Green, A., Brown, A., & Oram, R. N. (1980). *Determination of outcrossing rate in a breeding population of Lupinus albus L. (white lupin).*

Harrison, J. E. M. (1980). *The Genetic Contol of Alkaloid Content in Lupinus albus L., L. angustifolius L. and L. mutabilis Sweet*. University of Reading.

Harrison, J. E. M., & Williams, W. (1982). Genetical control of alkaloids in Lupinus albus. *Euphytica*, *31*(2), 357–364. https://doi.org/10.1007/BF00021651

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, *50*(3), 346–363.

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

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

Li, X., Yang, H., Buirchell, B., Yan, G., Li, X., Yang, H., Buirchell, B., & Yan, G. (2011). Development of a DNA marker tightly linked to low-alkaloid gene iucundus in narrow-leafed lupin (Lupinus angustifolius L.) for marker-assisted selection. *Crop and Pasture Science*, *62*(3), 218–224. https://doi.org/10.1071/CP10352

Lucas, M. M., Stoddard, F., Annicchiarico, P., Frias, J., Martinez-Villaluenga, C., Sussmann, D., Duranti, M., Seger, A., Zander, P., & Pueyo, J. (2015). The future of lupin as a protein crop in Europe. *Frontiers in Plant Science*, *6*. https://www.frontiersin.org/articles/10.3389/fpls.2015.00705

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

Martin, G. E., Rousseau-Gueutin, M., Cordonnier, S., Lima, O., Michon-Coudouel, S., Naquin, D., de Carvalho, J. F., Aïnouche, M., Salmon, A., & Aïnouche, A. (2014). The first complete chloroplast genome of the Genistoid legume Lupinus luteus: Evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Annals of Botany*, *113*(7), 1197–1210. https://doi.org/10.1093/aob/mcu050

Matsushima, R., Hu, Y., Toyoda, K., Sodmergen, & Sakamoto, W. (2008). The Model Plant Medicago truncatula Exhibits Biparental Plastid Inheritance. *Plant and Cell Physiology*, *49*(1), 81–91. https://doi.org/10.1093/pcp/pcm170

Otterbach, S. L., Yang, T., Kato, L., Janfelt, C., & Geu-Flores, F. (2019). Quinolizidine alkaloids are transported to seeds of bitter narrow-leafed lupin. *Journal of Experimental Botany*, *70*(20), 5799–5808. https://doi.org/10.1093/jxb/erz334

Park, H.-S., Lee, W. K., Lee, S.-C., Lee, H. O., Joh, H. J., Park, J. Y., Kim, S., Song, K., & Yang, T.-J. (2021). Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four

reciprocal F1 hybrid combinations. *Scientific Reports*, *11*, 2506. https://doi.org/10.1038/s41598-021-81988-w

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2021). *nlme: Linear and Nonlinear Mixed Effects Models*. https://CRAN.R-project.org/package=nlme

Plarre, W. (1985). Lupinen (Lupinus spec.). In *Lehrbuch der Züchtung landwirtschaftlicher Kulturpflanzen* (pp. 185–197).

Porsche, W. (1964). Untersuchungen über die Vererbung der Alkaloidarmut und die Variabilität des Restalkaloidgehaltes beiLupinus albus L. *Der Züchter*, *34*(6), 251–256. https://doi.org/10.1007/ BF00705826

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

Święcicki, W., Górny, A., Barzyk, P., Gawłowska, M., & Kaczmarek, Z. (2019). Genetic analysis of alkaloid accumulation in seeds of white lupin (Lupinus albus L.). *Genetika*, *51*(3), 961–973. https://doi.org/10.2298/GENSR1903961S

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

Wang, Q., Zhang, Y., Fang, Z., Liu, Y., Yang, L., & Zhuang, M. (2012). Chloroplast and mitochondrial SSR help to distinguish allo-cytoplasmic male sterile types in cabbage (Brassica oleracea L. var. Capitata). *Molecular Breeding*, *30*(2), 709–716. https://doi.org/10.1007/s11032-011-9656-9

Williams, W., Harrison, J. E. M., & Jayasekera, S. (1984). Genetical control of alkaloid production in Lupinus mutabilis and the effect of a mutant allele Mutal isolated following chemical mutagenesis. *Euphytica*, *33*(3), 811–817. https://doi.org/10.1007/BF00021907

Wink, M., & Witte, L. (1984a). Turnover and transport of quinolizidine alkaloids. Diurnal fluctuations of lupanine in the phloem sap, leaves and fruits of Lupinus albus L. *Planta*, *161*(6), 519–524. https://doi.org/10.1007/BF00407083

Wink, M., & Witte, L. (1984b). Turnover and transport of quinolizidine alkaloids. Diurnal fluctuations of lupanine in the phloem sap, leaves and fruits of Lupinus albus L. *Planta*, *161*(6), 519–524. https://doi.org/10.1007/BF00407083

Yoosefzadeh-Najafabadi, M., Hesami, M., & Rajcan, I. (2023). Unveiling the Mysteries of Non-Mendelian Heredity in Plant Breeding. *Plants*, *12*(10), 1956. https://doi.org/10.3390 /plants12101956

Chapter II: SNP marker validation and mapping of a new major QTL for low alkaloid content in white lupin

Abstract

White lupin is a protein crop with growing significance for protein self-sufficiency in Europe. There are white lupin cultivars that have been specifically bred to produce grains with minimised levels of toxic quinolizidine alkaloids (QAs). Yet, QA levels in grains from these "sweet" cultivars frequently exceed the recommended safety thresholds for use as food or feed, especially in response to abiotic stress. Hence, to facilitate pyramiding of multiple genes for low QA contents in white lupin breeding, this study aimed at validating molecular markers associated with this trait. The marker candidates were tested with a panel of 42 white lupin accessions and in segregating F2 populations. The marker candidates could not be confirmed to be associated with QA content. In addition, a bulk segregant analysis was conducted with bulks from phenotypic extremes of a F2 population, to map the sweetness gene from Dieta. A novel quantitative trait locus (QTL) for QA content in white lupin was detected on chromosome 5. To make this QTL amenable for marker assisted selection, further research and fine mapping is required.

Introduction

White lupin (*Lupinus albus* L.) is a grain legume with beneficial nutritional properties such as high content and quality of protein and oil, combined with nutritional fibre and the absence of starch. Furthermore, white lupins provide the ecosystem services of symbiotic nitrogen fixation, mobilisation of phosphate from the soil (Lucas et al., 2015, Talhinhas et al., 2016) and providing pollen to bees and bumblebees (Kleijn and Raemakers, 2008; Fijen et al. 2021). White lupin is more cold-tolerant but of similar yield potential and feed quality compared to soybean (Tröster et al., 2023), which is the most important protein source for feed in Europe, with 2.7 million t produced and 17.1 million t imported to in 2021 (FAO Stat). The role of white lupin as locally adapted alternative to soy is relevant, because there are extensive efforts of the EU to reduce the volume of soybean imports (Boerema et al., 2016; Hörtenhuber et al., 2011; Meul et al., 2012), since soy-imports from South America have been linked to rainforest destruction and the emission of greenhouse gasses (Sasu-Boakye et al., 2014; Barona et al., 2010).

However, lupin grains contain toxic and bitter quinolizidine alkaloids (QAs). Seeds of bitter white lupin cultivars contain about 2,000-4,000 mg kg^{-1} and seeds of extremely bitter wild material up to 127,000 mg kg^{-1} (Kroc et al., 2017). The recommended QA concentration thresholds are 200 mg kg-1 for human consumption and 500 mg kg^{-1} for animal feed (ANZFA, 2001). QAs are water soluble and can be removed by cooking and soaking (debittering) the grains for several days (Erbas 2010). This however is laborious, uses large quantities of water and reduces the feasibility of lupin use, especially for feeding purposes. Hence, low QA content (lQA) was historically, and continues to be, an important breeding target for lupins. There has been research on the genetics of QA content and breeding efforts in white lupins since the 1920s, resulting in "sweet" lQA cultivars (von Sengbusch, 1930 as cited by Li et al. 2011). Different individual genes are known to significantly reduce the QA contents in white lupins (lQA genes). The most effective and most commonly deployed lQA gene is *pauper*, which was characterised to cause a reduction of the seed QA contents in white lupin grains to 200-500 mg kg^{-1}. Other lQA genes were described to have a smaller effect, reducing the QA contents to ≤1,000 mg kg^{-1} (*mitis*, *reductus*, *exiguus*, *nutricius*). In addition, there are further lQA genes, which have been mentioned in literature, but without indication of the magnitude of their effects (*suavis*, *minutus*) (Harrison & Williams, 1982).

Despite the magnitude of the effect of the *pauper* gene in reducing the QA contents, grains harvested from lQA cultivars with *pauper* have been frequently observed to exceed the advised consumption thresholds. For example, eleven of twenty analysed seed batches of the anthracnose resistant cultivar Frieda had QA contents exceeding 500 mg kg^{-1}, with only two batches falling below 200mg kg^{-1} (FiBL data from 2020-2022, unpublished). Increased QA contents occur in response to drought and heat stress (Annicchiarico et al., 2014). The fact that "sweet" lQA cultivars with *pauper* do not reliably produce grains safe for food or feed, is a substantial economic risk for farmers. This issue might

aggravate in the future due to the climate crisis, with rising likelihoods of drought or heat stress (Calvin et al., 2023). To ensure economic feasibility of white lupin crops in future, it is desirable to breed for lupin cultivars with further reduced QA contents and reduced risk of exceeding consumption thresholds. One potential breeding approach for this target is to combine different lQA genes (Porsche, 1964).

The phenotypic difference between plants with two lQA genes and plants with one lQA gene may be relevant as pointed out above, but also likely relatively subtle. Selection of plants with the desired combination of lQA genes would require a high resolution in phenotyping: analytical quantification (*e.g.* GC-MS, HPLC-MS/MS, LC-MS/MS), which is destructive to the samples and costly. Conversely, if markers were available for multiple lQA genes, breeding could be facilitated by marker-assisted selection (MAS). Maintenance breeding may be facilitated likewise.

For *pauper*, the causal single nucleotide polymorphism (SNP) has recently been published. The SNP located on chromosome 18 at 12359687 bp with the lQA allele coded "TT" and hQA coded "TA" and "AA" on the antisense strand (3'-5'). The SNP impairs the function of an acetyltransferase involved in the early steps of the QA biosynthesis and hence the entire pathway (Mancinotti et al., 2023).

In an unpublished study (Schwertfirm et al., under review), a genome-wide association study (GWAS) and genomic prediction were conducted on a white lupin collection (LUW panel) that consisted of 170 gene bank accessions (IPK, Gatersleben, Germany), 16 cultivars and 73 elite lines from a breeding program of the Landwirtschaftliche Lehranstalten Triesdorf (LLT, Germany). The LUW panel served for marker development for QA content, anthracnose resistance and seed protein content. The panel comprised 164 accessions assumed to be "sweet", 3 accessions of unknown QA content and 93 assumed "bitter" content. 100 of the accessions, including the 73 breeding lines, were known to carry *pauper*. For 53 of the lQA accessions, the causal gene for sweetness was unknown. The alkaloid content was phenotyped by near-infrared spectroscopy. Those SNP markers, that are significantly associated with QA content and other than the causal mutation for *pauper*, are hence considered marker candidates for *non-pauper* lQA genes.

The aim of this study is the validation of these *non-pauper* marker candidates in a panel of 42 accessions (QA panel) and segregating F2 populations to facilitate pyramiding of multiple lQA genes in white lupin breeding. The QA panel was genotyped for 16 marker candidates. Subsequently, three segregating F2 populations were studied, which were segregating for more than one lQA gene. The populations were genotyped for *pauper* and those SNPs most likely to be polymorphic and segregating, based on the genotyping results of the parental genotypes in the QA panel. This was conducted to reveal whether the marker candidates explain those components of phenotypic variation, that remained after excluding all plants homozygous for *pauper*. Since most of the candidate SNPs are located on other chromosomes than chromosome 18, we hypothesised that these may be linked to *non-pauper* genes and hence may explain the mentioned phenotypic variation.

An important limitation of the LUW panel for the identification of markers associated with *non-pauper* lQA genes is that these genes are underrepresented. Two accessions were suspected to carry *suavis*, four accessions to carry *exiguus*, one to carry *nutricius*, and one (Dieta) to carry an unknown *non-pauper* gene. Two of the three populations were crossed to segregate for i.a. *exiguus*, which was the most represented *non-pauper* gene in the LUW panel. The third population was derived from a cross designed for segregation of the unknown lQA gene from Dieta and *pauper* (Dieta × Frieda). Dieta was generated by mutagenesis breeding in the Ukraine (Kurlovich 2002 as cited by Hufnagel 2021). Hence the unknown *non-pauper* lQA gene from Dieta is likely only present once in the LUW panel and therefore it is unlikely that one of the significant marker candidates is associated with this specific lQA gene. To map the unknown *non-pauper* gene from Dieta, we conducted a bulked segregant analysis (BSA) by bulking and sequencing eight plants from the respective phenotypic extremes, excluding all plants homozygous for *pauper*. Thereby we tested if there are genomic regions, which are monomorphic within, but polymorphic between the bulks. We hypothesised, that this should be the case and that this would allow to detect a major QTL related to alkaloid content.

Material and Methods

Plant material & QA phenotyping

A panel of 42 accessions was composed to comprise different lQA genes as well as the parentals and precursor F1 plants of the F2 populations that were the target of this study (Suppl. Table II - 1). All but three samples were taken from frozen leaf material, that had been collected during previous pot or field trials. The frozen leaf material, from which the DNA was later isolated, was phenotyped by sqeezing sap from the petioles or leaves onto a Dragendorff-Paper. Dragendorff-Paper is a filter paper soaked in iodine/potassium iodide solution (Dragendorff solution) and dried back at 40°C in a drying cabinet (Plarre 1985). A visible reddish-brown reaction on of the Dragendorf reagent was described to occur for QA contents of ≥ 500mg kg^{-1} (Li et al., 2011).

The F2 populations (Table 3) were chosen to have parentals with contrasting lQA genotypes according to literature, and hence to segregate for two lQA genes respectively. From the segregation of two complementary lQA genes in an F2 population, a phenotypic segregation ratio of 7:9 (lQA:hQA) is expected if both parentals were homozygous (Suppl. Table II - 2). The actual genotypes in accession seed lots often deviate from the assumed ones and high degrees of heterozygosity are commonly observed due to challenges in isolation from insect pollination during seed multiplication by genebanks (FiBL own, unpublished data). To ensure that segregation of more than one lQA gene was present in the F2 populations, the segregation patterns of QA contents were analysed and compared to the genotypes for the *pauper* causal SNP.

Table 3: F2 populations for segregation analysis with their crossing formula, lQA genes assumed to be segregating and the expected phenotypic segregation pattern. The population with grey shading was included to be segregating for pauper only, to serve as check.

FiBL ID	Crossing formula	lQA genes assumed to be crossed	Expected segregation
21.L.26	Boros × Nährquell	*pauper*[1] × *nutricius*[1]	**7:9**
21.L.35	Murringo × Butan	*pauper* × *exiguus*[1]	**7:9**
21.L.57	Neuland × Kiev mutant	*exiguus*[2] × *pauper*[1]	**7:9**
21.L.37	Breeding line × Nährquell	*pauper* × *nutricius*[1]	**7:9**
21.L.23	Neuland × Nährquell	*exiguus*[2] × *nutricius*[1]	**7:9**
21.L.97	Dieta × Frieda	*unknown non-pauper*[2] × *pauper*	**7:9**
21.L.51	VIR94 × Breeding line	*wildtype bitter* × *pauper*	**7:9**
21.L.73	Peragis Stamm × Frieda	*suavis*[3] × *pauper*	**7:9**
21.L.68	Frieda × Peragis Stamm	*pauper* × *suavis*[3]	**7:9**
21.L.44	Kiev mutant × Pop2078	*pauper*[1] × *bitter*	**1:3**

[1] Rychel and Ksiazkiewicz, 2019
[2] Lin et al., 2008
[3] Šatović, 1993

The F2 populations for this study were selected from the breeding program at FiBL. They were sown on the field in Leibstadt (47°35'26.4"N 8°10'56.0"E), in 10 row plots with 14 sown seeds per row for each population, flanked by single rows with six sown seeds for the parental genotypes respectively. The alkaloid content of the individual plants of the populations was phenotyped by squeezing liquid from the petioles on Dragendorff-paper at the onset of flowering (BBCH 63-65, 02/06/2023). The reaction was scored with 1 for a purely white stain, 2 for white stain with reddish encirceling, 3 for a pinkish stain with a lighter center and 4 for an intensely red stain without discoloration in the center Suppl. Figure II - 1). Additionally, bitterness was phenotyped by Christine Arncken (FiBL lupin breeder) by scoring the taste of one flower per plant in a range from 1 (absence of bitterness) to 9 (extremely bitter). Due to lacking phenotypic segregation, 7 of the originally 10 F2 populations were excluded. Leaf samples were collected and frozen for later DNA isolation from all individuals in the 3 focus populations and the parental individuals. The DNA was isolated from the frozen leaves with Whatman FTA cards.

SNP Marker candidates

In previous work by LfL in collaboration with FiBL, the white lupin LUW panel was used for identifying SNP markers associated with low-alkaloid content in the seeds via (i) GWAS and (ii) genomic prediction followed by "marker reduction".
Schwertfirm et al. (under review) conducted the GWAS based on adjusted means of QA contents (NIRS) correcting for the year effect (trials had been conducted in 2020 and 2021) with a total number of 166 samples. Initially, the analysis was conducted with SNPs from genotyping by sequencing (GBS) data complemented with 96 SNPs to enrich along the *pauper* locus (since the causal mutation hadn't been published at that time), using published variants in the white lupin genome browser (https://whitelupin.fr). Different models were used: GLM, MLM, MLMM, FarmCPU and Blink.
A second GWAS was performed including additional markers from targeted GBS for six loci associated with anthracnose resistance. In their publication (under review) Schwertfirm et al. will report only four markers from the second GWAS by MLMM, FarmCPU and BLINK (since these models account for the kinship structure in the LUW panel) and at least 90 genotyped plants, -log(p) > 5.7 and a minor allele frequency of >0.02. For marker validation in this thesis, all significant marker-trait associations with $p < 5.7$ were included (Table 4).
Additionally, Schwertfirm et al. performed genomic prediction using a ridge regression model followed by "marker-reduction" (not included in the publication). First, the genomic prediction was performed using all markers and all phenotyped accessions in the LUW panel as training set and the 39 re-sequenced accessions and respective genotypes from the public white lupin genome browser as test set. In a following step ("marker reduction") the genomic prediction was performed including the two markers in agreement with the GWAS on chromosome 18 (Lalb_Chr18_12463263 and Lalb_Chr18_12359687) plus the 10 markers with the highest effect (Table 5) in order to get model to get a classification accuracy of bitter/sweet of 85% (true to predicted).

Table 4: Causal SNP for *pauper* and candidate SNPs from GWAS with association to QA content in white lupin seeds, sorted by p-value. MAF is the minor allele frequency. "n" is the number of genotyped plants with this marker. "ref." and "alt." are the reference and alternative alleles, respecively. "X" represents a deletion. "N" represents an unknown nucleotide for the alternative allele.

chromosome	location	analysis	model	MAF	n	-log(p)	effect	ref.	alt.
Lalb_Chr18 [1,2]	12359687	GWAS2	MLM	0.48	162	8.33	-0.47	T	A
Lalb_Chr16 [2]	12866499	GWAS2	FarmCPU	0.06	149	7.01	0.61	T	N
Lalb_Chr16	2668267	GWAS1	FarmCPU	0.18	121	6.97	0.3	C	T
Lalb_Chr14 [2]	5832122	GWAS2	FarmCPU	0.3	91	6.75	-0.26	G	X
Lalb_Chr00c12	45679	GWAS1	FarmCPU	0.45	23	6.72	0.44	AC	TA
Lalb_Chr18 [2]	12463263	GWAS2	FarmCPU	0.34	154	6.58	0.21	A	G
Lalb_Chr18	12457069	GWAS1	MLMM	0.31	159	6.02	0.38	A	T
Lalb_Chr11	1128526	GWAS1	FarmCPU	0.43	27	5.79	0.32	T	C
Lalb_Chr06	13713695	GWAS1	FarmCPU	0.04	165	5.77	-0.29	A	G
Lalb_Chr25	11617387	GWAS1	FarmCPU	0.3	103	5.73	-0.18	A	C

[1]causal mutation for *pauper*

[2]will be published (Schwertfirm et al. under review).

Table 5: Candidate SNPs from marker reduction based on a genomic prediction model for QA content in white lupin seeds, sorted by magnitude of the marker effect.

chromosome	location	marker effect	reference allele	alternative allele
Lalb_Chr18 [1]	12463263	.291	A	G
Lalb_Chr25	1199129	- .160	T	G

Lalb_Chr18 [1,2]	12359687	.148	T	A
Lalb_Chr12	2626435	.122	C	T
Lalb_Chr21	12633426	- .110	C	A
Lalb_Chr21	12637918	- .110	T	C
Lalb_Chr21	12637843	.096	T	C
Lalb_Chr18	12110444	.072	A	G
Lalb_Chr10	16473189	.058	C	T
Lalb_Chr18	12109519	.033	T	C
Lalb_Chr21	12361245	- .021	A	T
Lalb_Chr21	12633496	- .010	T	A

[1] also present in GWAS
[2] causal mutation for *pauper,* also present in GWAS

KASP marker genotyping and genetic segregation analysis

Genotyping was conducted using PCR markers (KASP). The 42 genotypes in the QA panel were phenotypes with all 17 marker candidates, that were significantly associated with QA content (primer sequences in Suppl. Table II - 3). Markers Lalb_Chr00c12_45679, Lalb_Chr11_1128526 and Lalb_Chr18_12109519 were not genotyped, because the KASP reaction was not successful. Subsequently, the selected F2 populations were genotyped for the causal mutation of *pauper*. To ensure that indeed two different lQA genes were segregating in the F2 populations, the phenotypic segregation and *pauper* genotypes were graphically and statistically analysed by calculating χ^2 for the assumed segregation pattern of 7:9 (lQA:hQA) (Suppl. Table II - 2).

Only three populations with segregation patterns that implied segregation for lQA genes other than / additionally to *pauper*, were further genotyped. The genotyping was conducted population-wise for only those markers, which had been polymorphic for the parental genotypes in the QA panel.

To calculate the effect of *pauper* in the populations, a linear model was fitted with populations and the genotype for pauper as explanatory variables for the QA score (response variable).

Since not every candidate marker was genotyped for each population, the statistical analysis of candidate marker effects was conducted individually for each population. The genotype for *pauper* ("TT", "TA", "AA") was included as variable in the models, to account for the variance explained by *pauper* and to discriminate against markers linked with *pauper* but without explanatory power themselves. The model assumptions were checked with the package *DHARMa* (Hartig, 2021) and pairwise comparisons were made using the package *emmeans* (Lenth, 2023), respectively.

Pooling for bulked segregant analysis

To conduct bulked segregant analysis (BSA) in the Dieta × Frieda F2 population, bulked samples were curated. Two bulked samples represented the phenotypic extremes of the Dieta × Frieda subpopulation excluding all plants homozygous *pauper* ("TT"). The "sweet" bulk was selected from plants with QA scores ≤8.5. One plant fitting these criteria was excluded from the "sweet" bulk, since 40% of the seeds from this plant had a pinkish hue, which can be indicative for high QA contents in white lupin seeds (FiBL unpublished data), hence the "sweet" bulk consisted of eight plants. To be consistent with the number of plants per bulk, the bitter bulk was comprised by eight plants with QA scores ≥13. From all plants in this phenotypic class, plants with the highest share of seeds with a pinkish hue were favoured.

To improve confidence in the variance calling from the sequencing of the bulks from the F2, we also included samples from the parental genotypes for sequencing. Since no DNA samples had been taken from the true parentals of the cross and heterogeneity is often observed in seed batches of market cultivars, samples for the parentals were pooled from plantlets, grown from the same seed batch as the true parental plants had been sourced. For consistency, the parental pooled for Dieta and the parental pooled sample for Frieda comprised eight plants respectively.

Frozen leaflets from the Frieda × Dieta F2 population and fresh leaflets from Frieda and Dieta plants were collected into 2ml Eppendorf tubes, lyophilised and milled to powder by bead beating (Quiagen Tissue Lyzer II, 2 ceramic beads, 40seconds at 30Hz). From the milled powder, quantities of 2.12 mg (±0.06) were taken and put together for all samples of the respective pool. The four resulting pooled samples were mixed again by bead beating and DNA extraction according to the protocol in Suppl. Table II - 4. The DNA concentrations were checked by Nanodrop and Q-Bit and the quality was visually checked by gel-electrophoresis. The whole genome resequencing was at Novogene, UK with a sequencing depth of 15x.

Sequence Data Analysis and selective sweeps identification

The sequence data was processed and analysed in a modified version of the procedure published by Schneider et al. (2022). Quality checks were performed with the *fastqc* tool (Andrews 2010). The reads were then aligned to the reference genome using *bwa mem* (Li, 2013). With *sambamba* (Tarasov et al., 2015) duplicated reads were removed using the markdub-function. Furthermore, the reads were sorted by their mapped position and reads that had been mapped to multiple locations (*e.g.* secondary alignments) were discarded. The variant calling was conducted only on SNPs that had been previously reported on the white lupin genome browser (with 39 genotypes) (Hufnagel et al., 2021) and that were polymorphic between the parental bulk sequences. The rationale of this was to improve the confidence despite the low bulk sizes and read depth. SNPs were considered heterozygous in a bulk, when the allele frequency was between 0.2 and 0.8. All SNPs with heterozygosity within a parental bulk were excluded. Furthermore, SNPs with low quality were filtered out, with a minimal quality threshold of 25 for the whole read and 30 per base call. SNPs with read depths below 8 were also filtered out. The SNPs were then assigned to predicted genes based on their position in the gene or a maximum of 1000 bp up- or downstream of it as described by Schneider et al. (2022). For the resulting haplotypes, the allele frequencies were calculated for the "sweet" and for the bitter bulk.

For the statistical analysis, count numbers were calculated from the haplotype frequencies and read depths. The statistical analysis was conducted using the package *BayPass* (Gautier et al. 2015), which is a Bayesian approach to identify *i.e.* selective sweeps. Effective population size was set to 100, to reflect the size of the F2 Dieta × Frieda. From the p-values of each haplotype and their mapped positions, Manhattan plots were generated.

Results

QA panel phenotyping and genotyping

The KASP genotyping was successful for 17 of the 20 candidate markers, comprising the marker for the causal SNP of the *pauper* locus. For the SNPs markers, the KASP reaction did not work.

The working KAPS markers were polymorphic in the panel of 42 white lupin accessions. For each accession, there were assumptions on the phenotype and presence of lQA genes (Suppl. Table II - 1), based on descriptions and references in literature. These assumptions were partly confirmed and partly refuted by the Dragendorff phenotyping and genotyping results for *pauper* (Suppl. Table II - 5). Seventeen accessions were carrying the "sweet" reference allele of *pauper* in homozygous state ("TT") (Suppl. Table II - 5), including three accessions assumed not to carry *pauper* (Gyulatania, Kalina and a breeding line derived from Neutra). All plants homozygous for *pauper* were confirmed to be "sweet" according to the Dragendorff phenotyping. The remaining 25 accessions in the panel were either heterozygous ("TA") or homozygous for the bitter allele ("AA") at the *pauper* SNP. Fourteen of these accessions (not carrying *pauper*) were bitter, out of which three had been specifically selected to serve as bitter wildtype control. Six accessions with bitter phenotype had been assumed to be sweet and carry a lQA-gene. Five bitter accessions were F1 plants and heterozygous for *pauper* hence also hypothesised to be bitter.

In addition, there were eleven "sweet" plants that were not homozygous for the "sweet" *pauper* allele. These plants hence may carry a *non-pauper* lQA gene in homozygous state (*non-pauper* lQA group). We compared the genotypes of these non-pauper "sweet" plants against the genotypes of the three bitter (hQA group) wildtype plants (Table 6). By comparing the variant calls for these two groups of accessions for each marker individually, the markers were grouped into two categories:

a) Monomorphic within the hQA group and polymorphic between lQA and hQA group
b) Polymorphic within the hQA group

Table 6: Genotypes of 11 accessions (not homozygous for *pauper*) for 16 SNP marker candidates and the causal SNP for *pauper* (extract from QA panel). The colors represent the genotypes with blue for the reference allele, green for heterozygous state and red for the alternative allele. The accessions are grouped according to their QA phenotypes and their assumed genotypes. There are 3 accessions with sweet phenotypes and assumed to carry pauper, however the genotyping contradicted this assumption. The markers were grouped by according to their poly- / monomorphism within or between the groups.

Accession	**Phenotype**	**genotype group**	**assumed lQA gene**	Lalb_Chr06_13713695	Lalb_Chr10_16473189	Lalb_Chr18_12110444	Lalb_Chr21_12633426	Lalb_Chr21_12637843	Lalb_Chr21_12637918	Lalb_Chr25_11617387	Lalb_Chr12_2626435	Lalb_Chr14_5832122	Lalb_Chr16_12866499	Lalb_Chr16_2668267	Lalb_Chr18_12457069	Lalb_Chr18_12463263	Lalb_Chr21_12361245	Lalb_Chr25_1199129	Lalb_Chr22_2016762	***pauper:*** Lalb_Chr18_12359687
LUP 2079	hQA	wildtype	-	AA	TT	GG	AA	CC	CC	CC	TT	XX	TT	TT	TT	GG	AA	TT	GG	AA
Pop 2078				AA	TT	GG	AA	CC	CC	CC	TT	GG	CC	CC	AA	AA	TT	GG	GT	AA
Weibit				AA	TT	GG	AA	CC	CC	CC	CC	GG	CC	CC	AA	AA	TT	GG	GG	AA
Butan	lQA	non-pauper	exiguus	AA	TT	GG	CC	TT	TC	CC	CC	XX	TT	CC	AA	AA	AA	TT	GG	AA
Start			exiguus	AA	CT	GG	CC	TT	TC	CC	CC	XX	TT	TT	TT	GG	AA	TT	GG	AA
Dieta			unknown	AA	TT	GG	AA	CC	CC	CC	CC	GG	CC	TT	TT	GG	TT	TT	GT	AA
Nährquell			nutricius	AA	CC	GG	CC	TT	TC	CC	CC	XX	TT	CC	AA	AA	AA	TT	GG	AA
Nährquell			nutritius	AA	CC	GG	CC	TT	TT	CC	CC	XX	TT	CC	AA	AA	AA	TT	GG	AA
Energy			pauper (?)	AA	CT	AA	CC	TT	TT	AC	CC	XX	TT	CC	AA	AA	AA	TT	GG	TA
Neutra			pauper (?)	GG	CC	AA	CC	TT	TT	CC	CC	XX	TT	TT	TT	GG	AA	TT	GG	AA
Zulika			pauper (?)	AA	CC	AA	CA	TC	TC	CC	CC	GG	TT	CC	AA	AA	AT	TT	GT	AA
SNP marker grouped by occurrence:				a) monomorphic within hQA group and polymorphic between hQA and lQA							b) polymorphic within hQA									*pauper*

F2 QA phenotyping and segregation analysis

The phenotyping scores for the QA test in the F2 populations with Dragendorff paper are positively correlated with the phenotyping scores of the tasting (Spearman's rank correlation rho = 0.49, p-value < 2.2e-16; Suppl. Figure II - 2).

The F2 populations comprised 135 plants (Neuland × Kiev mutant), 110 plants (Neuland × Nährquell) and 113 plants (Dieta × Frieda). These populations segregated phenotypically with QA scores ranging from 4.5 to 15.

The three populations segregated for *pauper* (Lalb_Chr18_12359687), including Neuland × Nährquell for which both parentals were supposedly not carrying *pauper*. The baseline assumption for the segregation of *pauper* in the F2 populations of 1:2:1 could not be rejected (Table 7). The *pauper* genotype "TT" significantly reduced the QA score of plants in the populations (Table 8, Figure 3), whereas there were no significant differences between TT and TA (p=0.444, df=342, t-ratio=0.766, SE=0.314, estimate = 0.241). With *pauper* known to segregate in the three populations, the simplest assumption regarding the phenotypic segregation of the QA content in the F2 populations was 1:3 (for … vs. ….). This hypothesis could be rejected for all three populations. Assuming segregation of two independent recessive lQA genes, 7:9 segregation of the QA content could not rejected in any of the three populations (Table 9).

To further validate segregation of other lQA genes in the populations, the phenotypic segregation was investigated further in subpopulations excluding all plants homozygous for *pauper* ("TT") in comparison to the whole populations (Figure 4). With *pauper* excluded and 7:9 segregation not rejected in the F2s, the subpopulations were hypothesised to segregate with 1:3 ratio. This hypothesis was sustained for Neuland × Nährquell but rejected for Dieta × Frieda and Neuland × Kiev mutant (Table 9).

Table 7: Segregation ratios and χ^2 statistics for pauper in three F2 populations.

F2 population	expected segregation	observed segregation	df	χ^2	p-value
Neuland × Nährquell	1:2:1	31:57:22	2	1.618	0.4453
Neuland × Kiev mutant	1:2:1	32:67:36	2	0.244	0.885
Dieta × Frieda	1:2:1	31:47:34	2	3.054	0.2172

Table 8: Effect of the pauper allele "TT", compared to the other genotypes ("TA" / "AA") on the QA score in three segregating F2 populations, estimated as post-hoc analyses of a mixed model.

F2 population	estimated effect of TT on QA score against TA / AA	standard error	df	t ratio	p-value
Neuland × Nährquell	-4.31 / -4.61	0.53 / 0.66	342/335	8.07 / 6.97	<.0001/<.0001
Neuland × Kiev mutant	-1.91 / -2.44	0.51 / 0.59	339.9/342	3.718/4.142	0.0002/<.0001
Dieta × Frieda	-3.55 / -3.44	0.56 / 0.60	292.8/335.4	6.35 / 5.75	<.0001/<.0001

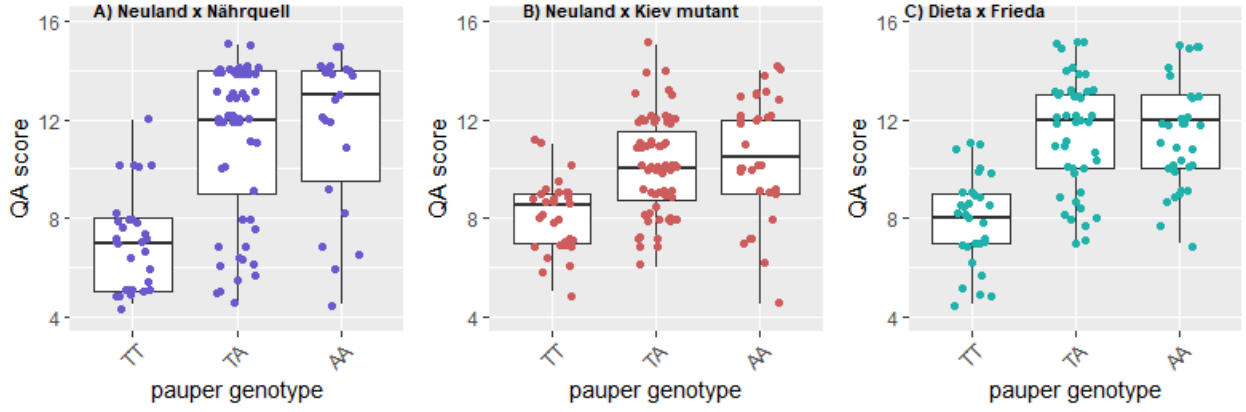


Figure 3: Genotype effects of *pauper* (Lalb_Chr18_12359687) on the QA score in three segregating F2 populations.

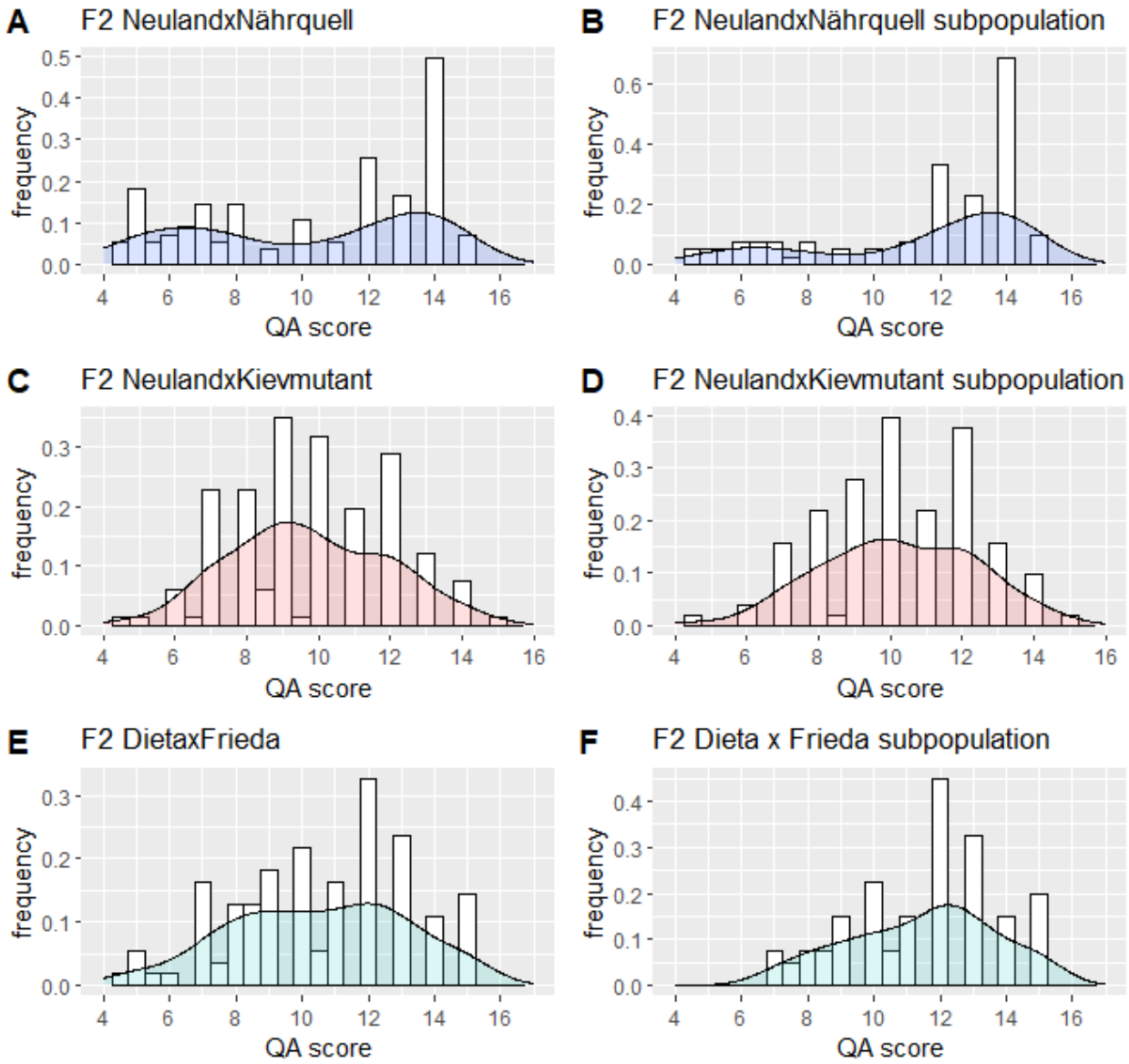


Figure 4: QA score frequencies and densities in F2 populations and subpopulations excluding individuals homozygous for *pauper* (Lalb_Chr18_12359687): A) F2 NeulandxNährquell B) F2 subpopulation NeulandxNährquell excluding pauper C) F2 NeulandxKievmutant D) F2 subpopulation Neulandx Kievmutant excluding pauper E) F2 DietaxFrieda F) F2 Dieta × Frieda subpopulation excluding pauper.

Table 9: Phenotypic segregation ratios and χ^2 statistics for F2 populations and their subpopulations excluding plants homozygous for *pauper*.

Segregating population		Expected ratio	Observed ratio	df	χ^2	p-value	
Neuland x Nährquell	whole population	1:3	46:63	1	17.202	<0.001	***
		7:9	46:63	1	0.14304	0.7053	
	subpopulation *pauper* excluded	1:3	21 :58	1	0.10549	0.7453	
Neuland x Kiev mutant	whole population	1:3	65:67	1	41.374	<0.001	***
		7:9	65:67	1	1.6181	0.234	
	subpopulation *pauper* excluded	1:3	37 :64	1	7.2904	0.0069	**
Dieta x Frieda	whole population	1 :3	41 :69	1	8.8364	0.00295	**
		7 :9	41 :69	1	1.8753	0.1709	
	subpopulation *pauper* excluded	1 :3	35 :45	1	15	<0.001	**

Candidate SNP markers

Six populations where excluded for further genotyping because of lacking phenotypic segregation in the Dragendorff test.

Population-wise genotyping for the markers which had been polymorphic for the parental genotypes in the QA panel included 5 SNPs for Dieta × Frieda, 6 for Neuland × Nährquell and 3 for Neuland × Kiev mutant. The reasons for the small number for Neuland × Kiev mutant is, that a) the crossed genotypes are genetically very distant (Hufnagel et al., 2021) and every marker was polymorphic for the parental genotypes in the QA panel (Suppl. Table II - 5), which didn't allow for prioritisation and b) that the segregation pattern did not suggest presence of one major QTL and to query for minor QTLs the phenotyping data would likely be too rough.

The genotypes of plants from seed lots of the cultivars used as parentals for the F2s showed genetic diversity contamination of the seed lots with other genotypes and heterozygosity (F2s Suppl. Table II - 5).

In the F2 from Dieta × Frieda, the markers Lalb_Chr25_11617387 and Lalb_Chr10_16473189 were monomorphic. The marker Lalb_Chr18_12457069 was polymorphic but had no significant effect on the QA content (Suppl. Table II - 7). The markers Lalb_Chr18_12463263 and Lalb_Chr16_2668267 co-segregated with *pauper*.

In the F2 from Neuland × Nährquell, the markers Lalb_Chr10_16473189, Lalb_Chr12_2626435, Lalb_Chr18_12463263, Lalb_Chr21_12637918 and Lalb_Chr25_11617387 were monomorphic. The marker Lalb_Chr14_5832122 was polymorphic, there was no significant effect on the QA content (Suppl. Table II - 7).

In the F2 population of Dieta × Kiev mutant, the marker Lalb_Chr14_5832122 was monomorphic. The marker Lalb_Chr25_11617387 was polymorphic, but not significantly associated with QA content. The marker Lalb_Chr18_12463263 co-segregated with *pauper*, with 4 recombinations in 127 plants.

In total, there were 5 recombinations between Lalb_Chr18_12463263 and Lalb_Chr18_12359687 (*pauper* causal mutation) in 232 plants (sum of successfully genotyped plants from the F2s of DietaxFrieda and NeulandxKiev mutant). The genetic distance is hence approximately 2.2 cM. There was no recombinant between Lalb_Chr16_2668267 and Lalb_Chr18_12359687 or Lalb_Chr18_12463263 in DxF with 105 plants, hence their genetic distance may be closer.

Bulked segregant analysis

There were 97 to 117 mio reads per bulk with an average read length of 150bp. The average coverage per base (read depth) ranged from 32 to 38 in the four bulks.

After quality filtering, mapping and removal of redundant 47-54 mio reads per bulk remained. Numbers of reads and reading depths per bulk in throughout this process is shown in Suppl. Table II - 8. There were 1,55 mio polymorphic SNPs (1,76 mio before q40 filtering) from the lupin browser that were polymorphic between the parents. After removal of SNPs heterozygous within the parental bulks and SNPs, poor quality (<25 whole read and <30 per base) and SNPs with read depth <8, the number of SNP was 514,982. The genotypes for these SNPs in the sweet and bitter bulks were used to construct 23,396 haplotypes, with an average of 21 SNPs per haplotype. The average read depth per haplotype was 298 in the sweet bulk and 304.797 in the bitter bulk. Plotting of p-values for the BayPass resulted in Figure 5. There is a pronounced peak in p-values on chromosome 5. The functional annotations of the genes associated with haplotypes with $\log_{10}$(p-value) > 4 are given in Suppl. Table II - 9. The predicted genes for these haplotypes are located in a region spanning from 5799140 bp to 7304977 bp. For the nine enzymes coded for, none could be directly linked to the QA biosynthetic pathway, for which only three enzymes in the first steps of the process are known: a lysine decarboxylase, a transaminase / copper amine oxidase and the acetyltransferase which is disrupted by the *pauper* gene. Furthermore, three transcription factors (TFs) of the families basic Helix-Loop-Helix, APELATA 2/ethylene-responsive element binding proteins and LIM.

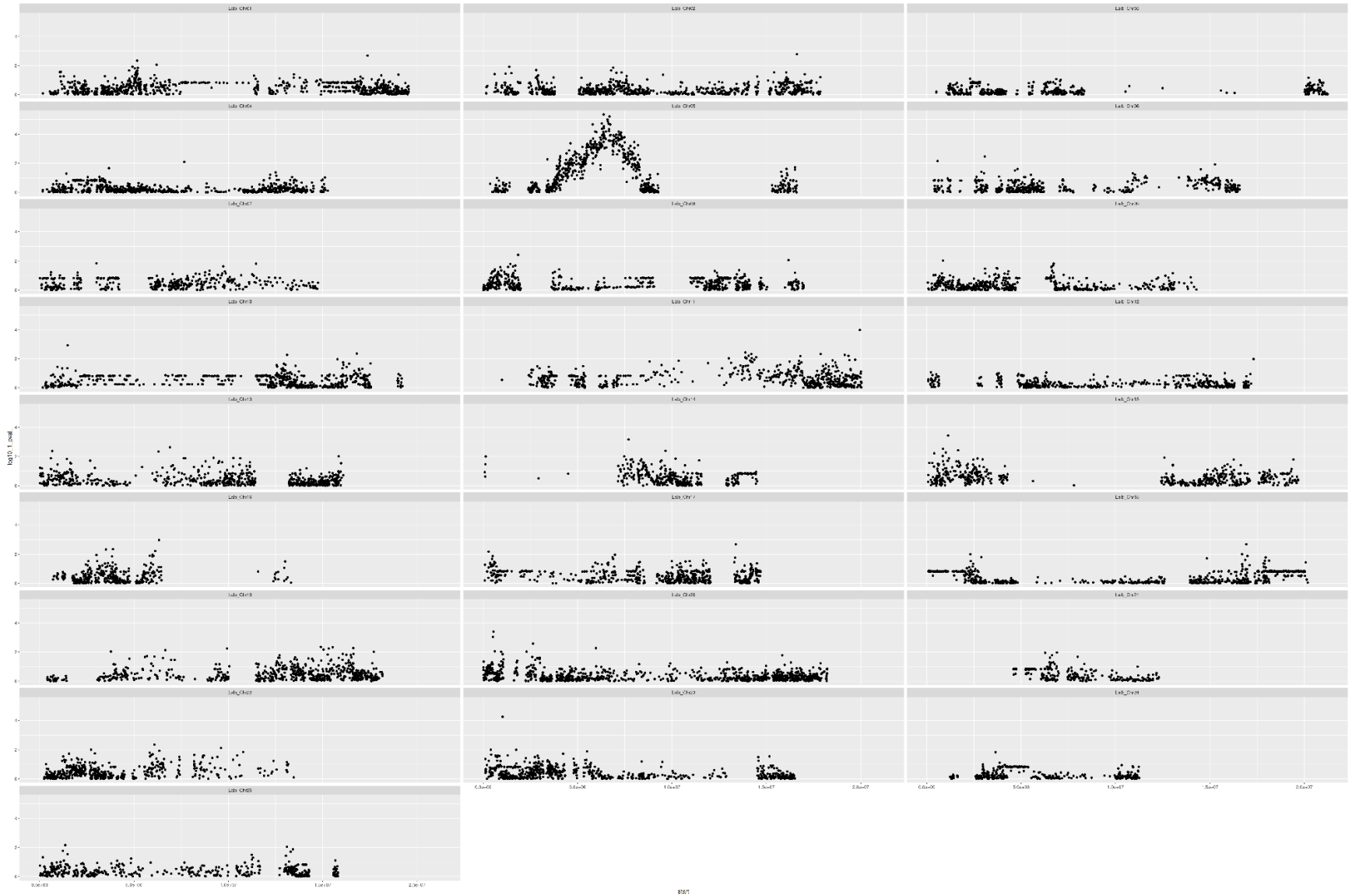


Figure 5: p values for differences in allele frequencies of haplotypes for two bulks from a white lupin F2 population, representing high and low QA levels. Each plot represents one chromosome, with the x-axis indicating the position of the haplotypes.

Discussion

We found, that by calculating a joint QA score from the two phenotyping methods taste scoring and Dragendorff test, the resulting data allowed to distinguish "sweet" from bitter plants, to analyse segregation patterns in F2 populations and to select plants from contrasting phenotypic extremes to successfully perform a BSA. The used methods for QA scoring were conducted on the field, provided immediate results, are non-destructive and very cost efficient in comparison to analytical quantification of QA contents. Since alkaloid levels in the plant tissues of lupins are fluctuating during the day (Wink & Witte 1984) and simultaneous testing of large populations is not possible, the confidence in combined phenotyping data may be higher due to the integration of two distinct measurements, rather than due to the particular combination of taste scoring with Dragendorff testing.

The observed segregation patterns in the F2 populations Dieta × Frieda and Neuland × Nährquell were indicative for segregation of two independent lQA genes, with no segregation distortion by *e.g*. natural selection or formation of deleterious gene combinations. This is consistent with earlier studies, where a 7:9 segregation pattern was observed (Harrison and Williams 1982; Porsche 1964). Since the Neuland × Nährquell F2 population was segregating for *pauper* although the parentals do not carry *pauper* and was monomorphic for markers that were polymorphic between the parental genotypes in the QA panel, it is doubtful that the crossed genotypes were actually Nährquell and Neuland. In this population however, a *non-pauper* gene is segregating and hence a BSA could be conducted, analogous to the BSA carried out in this study with the F2 derived from Dieta × Frieda. The subpopulation of Neuland × Kiev mutant may be segregating for two or more lQA genes.

From the segregating F2 populations, plants that are sweet and "AT" for *pauper* can be selected, as they are likely homozygous for a *non-pauper* sweetness gene but still segregating for *pauper*. In the F3, their single plant progenies will then segregate for *pauper*. By genotyping, single plants homozygous for pauper ("TT") may be selected, in which both sweetness genes are hence fixed. Therefore, it is possible to select plants homozygous for two sweetness genes with markers for only one of these genes. However, lupin breeding involves other breeding targets as well, *i.e.* introgression of resistance against anthracnose disease, selection for early maturity and protein yield (Talhinhas et al., 2016; Rychel et al., 2019; Annicchiarico et al. 2023). For maintenance and introgression breeding, MAS would be a facilitation.

The genotyping in the QA panel revealed the absence of *pauper* in a number of "sweet" white lupin accessions, which therefore can be considered sources for *non-pauper* lQA genes. The marker candidates for *non-pauper* lQA genes could, based on the QA panel, neither be confirmed nor rejected. The group of marker candidates monomorphic within the hQA group and polymorphic between the hQA and lQA accessions represent candidates for further inquiry and may even be causal SNPs. Multiple markers on chromosome 21 were part of this group, which might indicate presence of a QTL in this region. Further, the markers that were polymorphic within the hQA group may still be linked to lQA genes, but not diagnostic *per se* due to decayed linkage disequilibrium (Alkemade et al. 2022; Hufnagel et al. 2021). Unfortunately, three marker candidates (Lalb_Chr00c12_45679, Lalb_Chr11_1128526, Lalb_Chr18_12109519) did not work in the KASP assay. To validate these in future research, a different genotyping method may be tried, or alternatively, linked flanking SNPs may be used in a KASP assay.

For none of the candidate markers, an association with QA content could be confirmed in the investigated F2 populations. The GWAS / GS that had been conducted to identify these marker candidates had some limitations (Schwertfirm et al., under review). On the one hand, "sweet" accessions known to carry *non-pauper* lQA genes were underrepresented in the LUW panel and on the other hand, the density of markers used was relatively low given the high level of LD decay in white lupin (Schwertfirm et al., under review, Alkemade et al., 2022, Hufnagel et al., 2021). The observed co-segregation of *pauper* with a marker mapped on chromosome 16 could be due to an erroneous physical mapping of the corresponding DNA fragment.

The analysis of haplotype frequencies from sequencing of four bulked samples, revealed a distinct peak on Lalb_chr5, which putatively represents the location of the unknown *non-pauper* lQA gene

from the cultivar Dieta. The relatively high genetic distance between Dieta and Frieda resulted in a high density of polymorphic haplotypes. Furthermore, there were many recombinations in and around this region, which resulted in a bell-like shape of the QTL peak. The BSA with sequencing was an appropriate method to identify a new QTL for QA content in white lupin and it may be applicable to other monogenic traits for which phenotypes in segregating populations can be discriminated into distinct groups. The signal on chromosome 5 had not been picked up by the previous GWAS and genomic prediction, which corroborates the aforementioned limitations of the LUW panel to identify markers associated with *non-pauper* lQA genes.

For application in breeding however, the QTL region is still genetically too large to allow selection of the QTL with flanking markers. Due to the high degree of LD decay (<4 kb) in white lupin (Hufnagel et al., 2021), it is a challenge to find markers that would be polymorphic and therefore diagnostic in many different genetic backgrounds. Therefore, finer mapping of the lQA gene on Lalb_chr5 would be desirable to facilitate breeding by MAS.

The functional annotations of genes in white lupin are largely predicted based on model organisms that do not produce QAs (Hufnagel et al., 2020). Furthermore, the understanding of the biosynthesis of QAs and the involved enzymes is limited (Mancinotti et al. 2023), challenging the approach to narrow down candidate genes based their annotations. Hence it was not possible to link annotated proteins or enzymes to QA synthesis. However, it is known that QA biosynthesis occurs in response to biotic and abiotic stress and hence there must be regulatory pathways mediating this response. Interestingly, the APETALA2/ethylene response element binding proteins (AP2/EREBP) are TFs involved in the regulation of "stress, redox, hormonal and metabolite-related expressional control" (Dietz et al. 2010) and can be subsumed to the APETALA2/ethylene response (AP2/ERF) family (Xie et al. 2019). The AP2/ERF family was strongly suggested to be involved in regulating the QA biosynthesis in narrow leaved lupin (*Lupinus angustifolius*), as gene coding for an AP2/ERF TF is a candidate gene for the lQA gene *iucundus*, which is the most common lQA gene in narrow leaved lupin (Kroc et al. 2019). Moreover, AP2/ERF and Basic Helix-loop-Helix (bHLH) were shown to be involved in alkaloid biosynthesis in *Nicotiniana tabacum*, *Catharanthus roseus* & *Camellia japonica* and to respond to jasmonic acid (JA). JA is a plant hormone that mediates the response to abiotic stress and herbivorous attacks. JA may also be involved in regulation of QA biosynthesis in lupins, since QA levels in lupin plant tissues rise after wounding (Frick et al. 2017). Hence, the genes on Lalb_chr5 which were predicted to code for bHLH TF and AP2/EREBP are candidate genes for the location of the causal mutation for the "sweetness" QTL in Dieta.

To fine map the QTL on chromosome 5, the seeds harvested from the plants of the Dieta × Frieda F2 population can be sown, thanks to the non-destructive phenotyping methods used. The *pauper* genotyping data can be used to compose the mapping population from plants, that were homozygous for the bitter allele of *pauper*. To reduce phenotyping efforts, progenies from "sweet" F2 plants may be sown separately from progenies from bitter F2 plants, since the former will likely be uniformly sweet whilst the latter will segregate and hence require phenotyping. The genotyping could be conducted again by bulk sequencing. Since in the F3 the number of recombinations is doubled compared to a F2, a higher mapping resolution may be obtained by analysing at least the same number of plants per phenotypic extreme, and with the more plants, the higher the resolution.

Furthermore, a chemotype based approach could be performed to narrow down the gene candidates in the QTL region on chr. 5 as was conducted by Mancinotti et al. (2023). This would however only work, if the QTL impacted a step during the biosynthesis. Additionally, the effect of JA on QA levels in Dieta could be compared to the response of another "sweet" white lupin genotype carrying *pauper*. Since *pauper* is known to reduce QA contents by obstructing the early steps of QA biosynthesis (Mancinotti et al. 2023), the genotypes might respond differently to JA, if the lQA gene in Dieta would indeed be connected to a JA-responsive TF. If the lQA gene on Lalb_chr5 was disruptive to QA expression in response to stress, this gene might be well suited to stabilise QA contents in white lupin genotypes with *pauper* in response to stress. Alltogether, these findings represent an important step in the development of white lupins that reliably produce grains safe for use, also under stressful conditions.

Conclusion

The marker candidates investigated in this study could not be confirmed to be associated with QA content. However, with the current availability of a SNP marker for *pauper*, it is possible to select for plants carrying two lQA genes from segregating populations, by combined phenotypic and genotypic selection and subsequent genotypic selection in the next generation. It is still desirable to develop markers for *non-pauper* lQA genes to facilitate white lupin breeding. By conducting a BSA with sequencing in a segregating F2 population, we identified a new QTL for QA content in white lupin on chromosome 5. Within this QTL, there are two genes coding for TF families, that had been linked to the regulation of QA production, potentially in response to biotic and abiotic stress (Frick et al. 2017). For application of MAS on this lQA QTL, it is required to perform fine mapping. In conclusion, this study contributes to the understanding of QA genetics in white lupin and to mapping of *non-pauper* genes. These are important steps towards the development of white lupin cultivars with low QA contents under stress conditions, which becomes increasingly relevant in the face of the climate crisis.

Literature

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. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *135*(3), 1011–1024. https://doi.org/10.1007/s00122-021-04014-7

Annicchiarico, P., de Buck, A. J., Vlachostergios, D. N., Heupink, D., Koskosidis, A., Nazzicari, N., & Crosta, M. (2023). White Lupin Adaptation to Moderately Calcareous Soils: Phenotypic Variation and Genome-Enabled Prediction. *Plants*, *12*(5), Article 5. https://doi.org/10.3390/plants12051139

Annicchiarico, P., Manunza, P., Arnoldi, A., & Boschin, G. (2014). Quality of Lupinus albus L. (White Lupin) Seed: Extent of Genotypic and Environmental Effects. *Journal of Agricultural and Food Chemistry*, *62*(28), 6539–6545. https://doi.org/10.1021/jf405615k

ANZFA. (2001). *Lupin Alkaloids in Food A Toxicological Review and Risk Assessment*. Australia New Zealand Food Authority. https://www.foodstandards.gov.au/publications/documents/TR3.pdf

Barona, E., Ramankutty, N., Hyman, G., & Coomes, O. T. (2010). The role of pasture and soybean in deforestation of the Brazilian Amazon. *Environmental Research Letters*, *5*(2), 024002. https://doi.org/10.1088/1748-9326/5/2/024002

Boerema, A., Peeters, A., Swolfs, S., Vandevenne, F., Jacobs, S., Staes, J., & Meire, P. (2016). Soybean Trade: Balancing Environmental and Socio-Economic Impacts of an Intercontinental Market. *PLOS ONE*, *11*(5), e0155222. https://doi.org/10.1371/journal.pone.0155222

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W. L., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., … Péan, C. (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* (First). Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

Dietz, K.-J., Vogel, M. O., & Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma*, *245*(1–4), 3–14. https://doi.org/10.1007/s00709-010-0142-8

Erbas, M. (2010). The Effects of Different Debittering Methods on the Production of Lupin Bean Snack from Bitter Lupinus Albus L. Seeds. *Journal of Food Quality*, *33*(6), 742–757. https://doi.org/10.1111/j.1745-4557.2010.00347.x

FAOSTAT. (n.d.). Food and Agriculture Organization of the United Nations. Retrieved 3 October 2023, from https://www.fao.org/faostat/en/#data/QCL

FastQC A Quality Control tool for High Throughput Sequence Data. (2010). Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Fijen, T. P. M., Morra, E., & Kleijn, D. (2021). Pollination increases white and narrow-leaved lupin protein yields but not all crop visitors contribute to pollination. *Agriculture, Ecosystems & Environment*, *313*, 107386. https://doi.org/10.1016/j.agee.2021.107386

Frick, K. M., Kamphuis, L. G., Siddique, K. H. M., Singh, K. B., & Foley, R. C. (2017). Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/articles/10.3389/fpls.2017.00087

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

Harrison, J. E. M., & Williams, W. (1982). Genetical control of alkaloids in Lupinus albus. *Euphytica*, *31*(2), 357–364. https://doi.org/10.1007/BF00021651

Hartig, F. (2022). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (R package version 0.4.6) [Computer software]. http://florianhartig.github.io/DHARMa/

Hörtenhuber, S. J., Lindenthal, T., & Zollitsch, W. (2011). Reduction of greenhouse gas emissions from feed supply chains by utilizing regionally produced protein sources: The case of Austrian dairy production. *Journal of the Science of Food and Agriculture*, *91*(6), 1118–1127. https://doi.org/10.1002/jsfa.4293

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

Kleijn, D., & Raemakers, I. (2008). A Retrospective Analysis of Pollen Host Plant Use by Stable and Declining Bumble Bee Species. *Ecology*, *89*(7), 1811–1823. https://doi.org/10.1890/07-1275.1

Kroc, M., Koczyk, G., Kamel, K. A., Czepiel, K., Fedorowicz-Strońska, O., Krajewski, P., Kosińska, J., Podkowiński, J., Wilczura, P., & Święcicki, W. (2019). Transcriptome-derived investigation of biosynthesis of quinolizidine alkaloids in narrow-leafed lupin (Lupinus angustifolius L.) highlights candidate genes linked to iucundus locus. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-018-37701-5

Kurlovich, B. S. (2002). *Lupins: Geography, Classification, Genetic Resources and Breeding*. Bogouslav Kourlovitch.

Lenth, R. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (R package version 1.8.9) [Computer software]. https://CRAN.R-project.org/package=emmeans

Li, H. (2013a). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. https://doi.org/10.48550/ARXIV.1303.3997

Li, X., Yang, H., Buirchell, B., Yan, G., Li, X., Yang, H., Buirchell, B., & Yan, G. (2011). Development of a DNA marker tightly linked to low-alkaloid gene iucundus in narrow-leafed lupin (Lupinus angustifolius L.) for marker-assisted selection. *Crop and Pasture Science*, *62*(3), 218–224. https://doi.org/10.1071/CP10352

Lin, R., Renshaw, D., Luckett, D., Clements, J., Yan, G., Adhikari, K., Buirchell, B., Sweetingham, M., & Yang, H. (2009a). 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

Lucas, M. M., Stoddard, F., Annicchiarico, P., Frias, J., Martinez-Villaluenga, C., Sussmann, D., Duranti, M., Seger, A., Zander, P., & Pueyo, J. (2015). The future of lupin as a protein crop in Europe. *Frontiers in Plant Science*, *6*. https://www.frontiersin.org/articles/10.3389/fpls.2015.00705

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

Mancinotti, D., Michiko Frick, K., & Geu-Flores, F. (2022). Biosynthesis of quinolizidine alkaloids in lupins: Mechanistic considerations and prospects for pathway elucidation. *Natural Product Reports*, *39*(7), 1423–1437. https://doi.org/10.1039/D1NP00069A

Meul, M., Ginneberge, C., Van Middelaar, C. E., de Boer, I. J. M., Fremaut, D., & Haesaert, G. (2012). Carbon footprint of five pig diets using three land use change accounting methods. *Livestock Science*, *149*(3), 215–223. https://doi.org/10.1016/j.livsci.2012.07.012

Porsche, W. (1964). Untersuchungen über die Vererbung der Alkaloidarmut und die Variabilität des Restalkaloidgehaltes beiLupinus albus L. *Der Züchter*, *34*(6), 251–256. https://doi.org/10.1007/BF00705826

Riechmann, J. L., & Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. *Biological Chemistry*, *379*(6), 633–646. https://doi.org/10.1515/bchm.1998.379.6.633

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

Sasu-Boakye, Y., Cederberg, C., & Wirsenius, S. (2014). Localising livestock protein feed production and the impact on land use and greenhouse gas emissions. *Animal*, *8*(8), 1339–1348. Scopus. https://doi.org/10.1017/S1751731114001293

Šatović, Z. (1993). Biochemical and genetic characteristics of white lupine (Lupinus albus) Alkaloids. *Sjemenarstvo*, *10*(3–4), 139–154.

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), Article 1. https://doi.org/10.1186/s13007-022-00852-8

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Tröster, M. F., Heinz, M., & Durst, L. (2023). Determining the Value of Novel Feedstuffs in Imperfect Markets, Taking Lupinus albus as an Example. *Agriculture*, *13*(4), Article 4. https://doi.org/10.3390/agriculture13040867

von Sengbusch, R. (1930). Bitterstoffarme Lupinen. *Der Züchter*, *2*(1), 1–11. https://doi.org/10.1007/BF01812763

Wink, M., & Witte, L. (1984). Turnover and transport of quinolizidine alkaloids. Diurnal fluctuations of lupanine in the phloem sap, leaves and fruits of Lupinus albus L. *Planta*, *161*(6), 519–524. https://doi.org/10.1007/BF00407083

Xie, Z., Nolan, T. M., Jiang, H., & Yin, Y. (2019). AP2/ERF Transcription Factor Regulatory Networks in Hormone and Abiotic Stress Responses in Arabidopsis. *Frontiers in Plant Science*, *10*, 228. https://doi.org/10.3389/fpls.2019.00228

General conclusion and outlook

Altogether this MSc thesis highlights the complexity of quinolizidine (QA) genetics in white lupin.

The role of maternal inheritance was investigated studied by determining QA contents of selfed and crossed seeds grown on the same mother plant genotype. Furthermore, based on QA levels in F1 plants from reciprocal crosses, effects of the cytoplasm were investigated. The main findings (of Chapter I) and implications for future research and breeding are:

- Confirmation that QA content in seeds is determined by the maternal genotype. This means for breeding that the QA phenotypes of seeds reflect the genotypes of the previous generation. In addition, isolation is required in the seed production of white lupins to prevent unintentional outcrossing, but not in cultivation as feed or food.
- Significant differences regarding the QA content in reciprocal F1 plants suggest an effect of cytoplasmic variation. Because of this result, reciprocal crosses can be advisable in breeding programs to ensure that the full range of genetic variation is revealed and that selection can draw from the full genetic potential.
- However, these results and previous studies are inconclusive on the cause for different QA levels in reciprocal F1s. Further research is needed to elucidate the role of cytoplasmic variation for white lupin QA genetics and breeding.

Furthermore, this MSc thesis (Chapter II) aimed to validate markers for major determinants of QA levels in white lupin that are complementary to *pauper*, the most common gene for low QA contents in market varieties, for which the causal mutation was published 2023. We found that:

- Candidate SNP markers could neither be validated in a 42 accessions panel (including cultivars, gene bank accessions and breeding material) nor in three segregating F2 populations.
- Nevertheless, with the available marker for *pauper* combined with phenotyping it is possible to select plants homozygous for two genes for low QA contents, by selecting plants with low QA contents that are heterozygous for *pauper* and in their progeny for plants homozygous for pauper.

Additionally, a bulk segregant analysis was conducted on an F2 population, by sequencing bulks of F2 plants with contrasting QA levels, as well as bulks from the same cultivars as the parentals of the F2 (Chapter II), with the following results:

- Evidence for a QTL for QA contents in white lupin on chromosome 5. In this QTL, there are candidate genes coding for transcription factor families that had previously been linked to QA biosynthesis in lupins.
- Further research is required to fine map the novel QTL on chromosome 5 and make it useful for marker assisted selection in white lupin breeding.

This complexity represents a challenge but importantly also an opportunity, if the genetic determinants can be harnessed for white lupin breeding and development of cultivars with minimised and stable QA levels.

Acknowledgments

Firstly, I'd like to express my gratitude to Dr. Mariateresa Lazzaro for the excellent guidance, support, as well as her expertise during this time. Furthermore, I am thankful to Dr. Chris Maliepaard for his supervision with all the advice, useful feedback and insightful discussions.

I would like to thank András Patyi, who was my practical supervisor in the lab, a great supporter and who introduced me to the fascination of molecular biology. Many thanks to Christine Arncken for providing the populations and accessions for the crossings, for the planning of the field and greenhouse and for sharing her wealth of knowledge.

I'm grateful to Dr. Elisa Biazzi for meticulously conducting the analytical quantification of alkaloids in seeds and plant tissues. Many thanks also to Dr. Michael Schneider for carrying out the bioinformatics part of the BSA and patiently answering my questions. Many thanks to Dr. Grit Schwertfirm (LfL) for providing her previous work and for performing the crossing control.

Thanks to Adrien for continuing with the crosses and keeping the plants alive. Thanks to Arthur for protecting the plants from the thrips. Thanks to Johanna for supporting with the plants in the climate chamber. I'm grateful to to Tobi and Felix for proofreading parts of the manuscript. And lastly, I want to express my gratitude to Knut for being ever supportive and simply wonderful.

This thesis was conducted in the scope of the research project DIVINFOOD (divinfood.eu, European Union's Horizon 2020 research and innovation programme, Grant Agreement N°101000383).

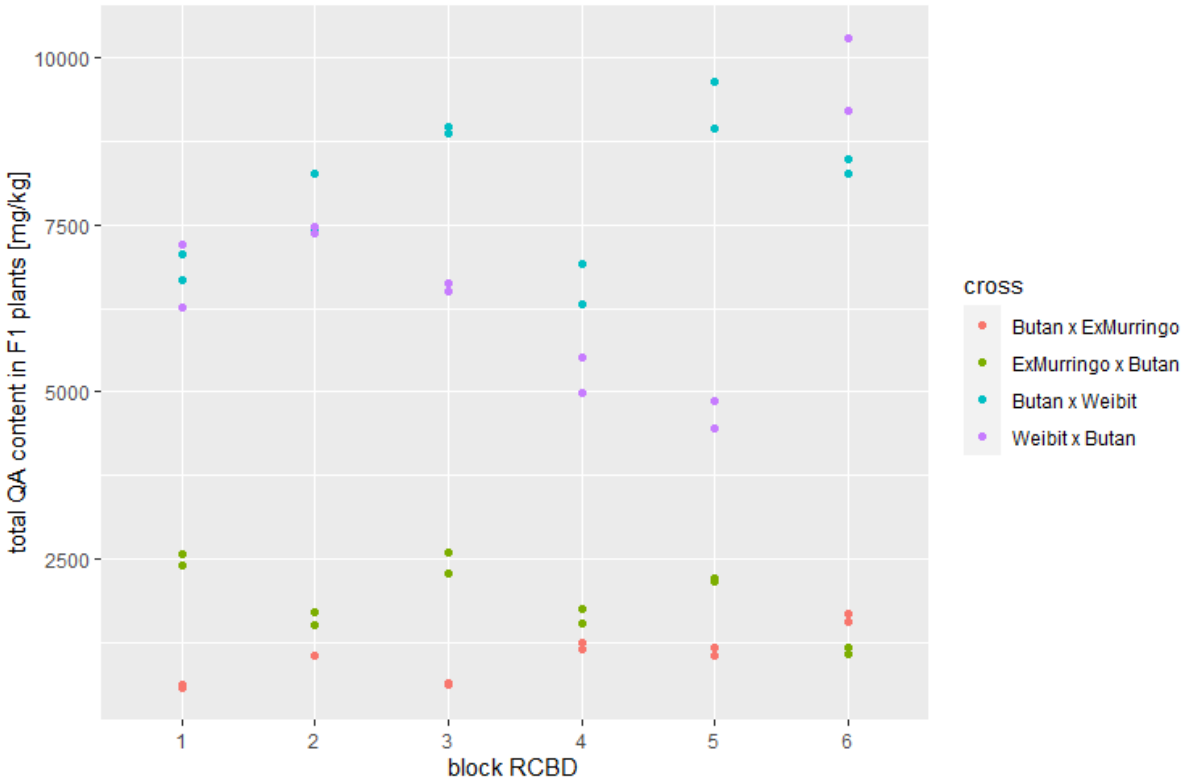
Appendices

Appendix Chapter I

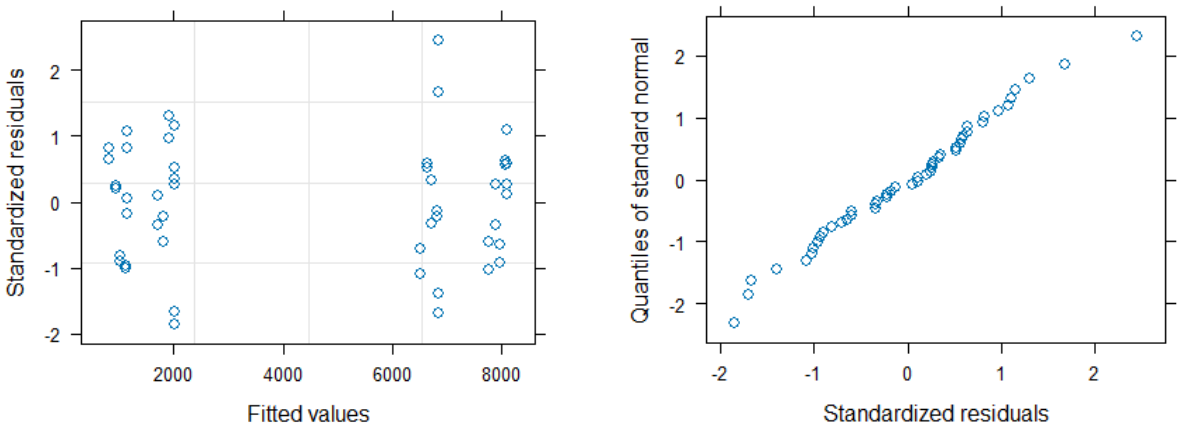
Suppl. Table I - 1: Composition of replicates of F1 seeds for analytical quantification of QAs. Each replicate consists of two seeds F1 seeds and, where possible, selfed seeds harvested from the same motherplant (P1).

Cross	P1 ID	number of F1 seeds from P1	number of selfed seeds from P1	Replicate number
Butan x ExMurringo	37	1	4	1
Butan x ExMurringo	38	1	4	1
Butan x ExMurringo	75	2	1	2
Butan x ExMurringo	79	1	4	2
Butan x ExMurringo	120	2	3	3
Butan x Weibit	73	2	1	1
Butan x Weibit	81	1	4	2
Butan x Weibit	84	1	4	1
Butan x Weibit	117	1	4	2
Butan x Weibit	119	1	2	3
Butan x Weibit	123	4	3	3
ExMurringo x Weibit	60	1	4	1
ExMurringo x Weibit	101	2	none[1]	2
ExMurringo x Weibit	102	1	none[1]	1
ExMurringo x Weibit	314	3	none[1]	3
ExMurringo x Butan	66	2	none[1]	1
ExMurringo x Butan	72	2	none[1]	2
ExMurringo x Butan	112	2	1	3
Weibit x ExMurringo	298	1	1	1
Weibit x ExMurringo	58	3	2	1
Weibit x ExMurringo	100	6	none[1]	2
Weibit x ExMurringo	27	2	none[1]	3
Weibit x Butan	87	4	2	1
Weibit x Butan	30	2	2	2
Weibit x Butan	49	3	3	3

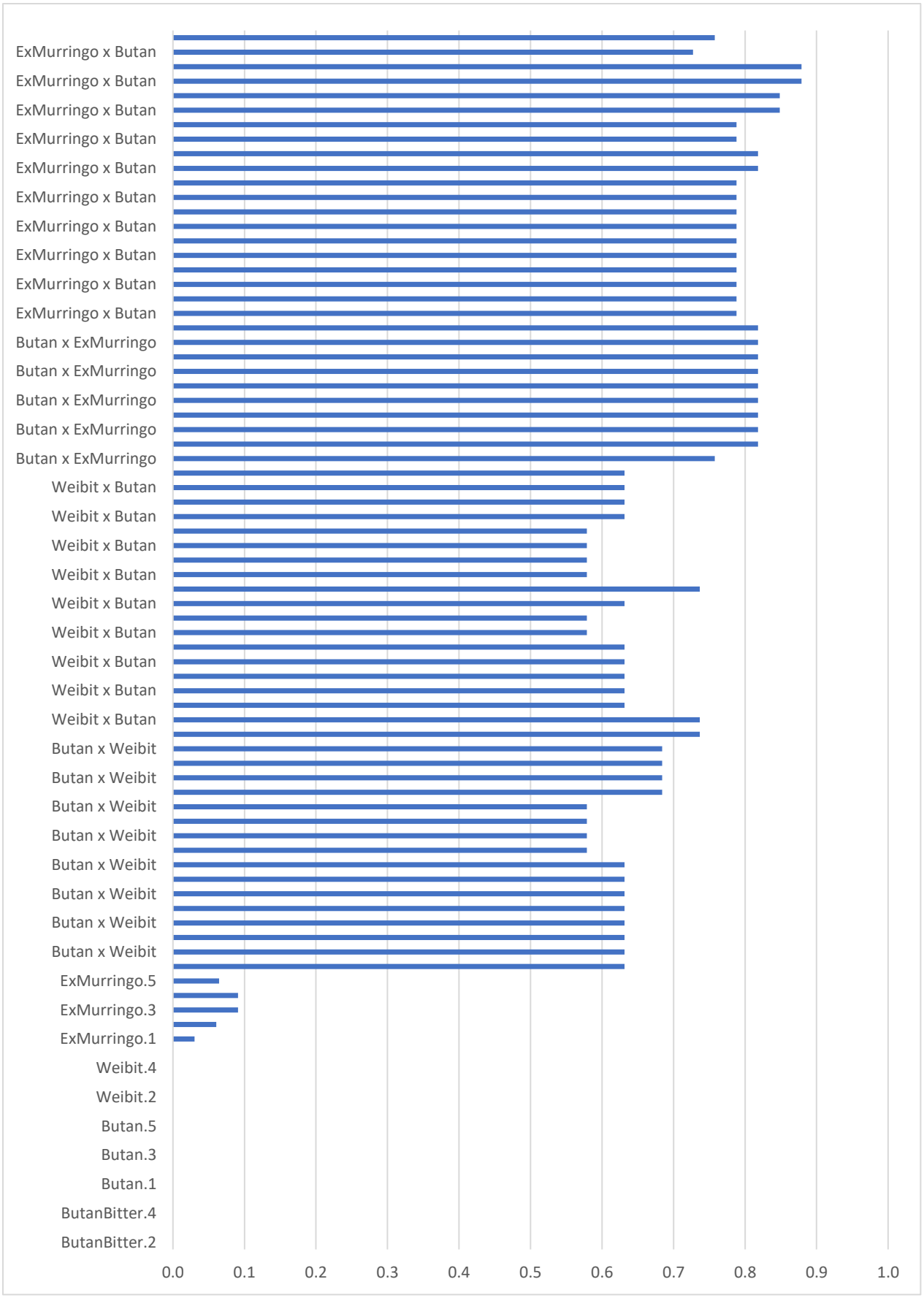
[1] Since no selfed seeds were available from these plants, selfed seeds from other plants of the same genotype were used for



Suppl. Figure I - 1: Total QA contents in F1 plants by crossing and block effects (RCBD).



Suppl. Figure I - 2: Residual plots for the statistical analysis QA content in reciprocal F1 plants by *gls()* from the package *nlme*.



Suppl. Figure I - 3: Rate of heterozygosity for F1 plants from genotyping with 96 SNP markers. There are two bars per sample because two 2 technical replicates per DNA sample.

Appendix Chapter II

Suppl. Table II - 1: Accessions in QA panel for marker validation, genotyping and checks for segregation analysis of F2 populations.

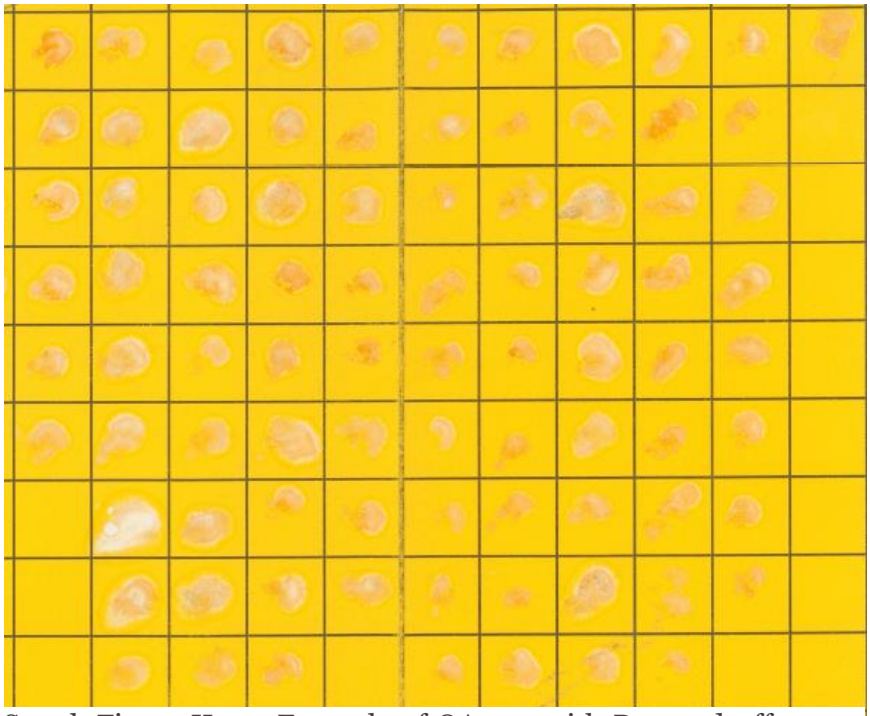
Accession name	**expected phenotype**	**Accession type**	**assumed sweetness gene**	**reason for inclusion into panel**
Weibit	bitter	cultivar	*wildtype*	marker validation bitter control
LUP 2079	bitter	genetic resource	*wildtype*	marker validation bitter control
Pop 2078	bitter	genetic resource	*wildtype*	marker validation bitter control
Butan	sweet	cultivar	*exiguus*	marker validation non-pauper
Start	sweet	cultivar	*exiguus*	marker validation non-pauper
Butan	sweet	cultivar	*exiguus*	marker validation non-pauper
Gyulatanya	sweet	cultivar	*exiguus (?)*	marker validation non-pauper
Kalina	sweet	cultivar	*exiguus (?)*	marker validation non-pauper
Neuland*	sweet	cultivar	*exiguus*	marker validation non-pauper
elatus aus Neutra	sweet	cultivar?	*non-pauper (?)*	marker validation non-pauper
Nährquell *	sweet	cultivar	*nutricius*	marker validation non-pauper
Peragis Stamm 749	sweet	breeding line	*suavis (?)*	marker validation non-pauper
Amiga	sweet	cultivar	*pauper*	marker validation pauper
Energy	sweet	cultivar	*pauper*	marker validation pauper
Feodora	sweet	cultivar	*pauper*	marker validation pauper
Frieda	sweet	cultivar	*pauper*	marker validation pauper
Kiev mutant *	sweet	cultivar?	*pauper*	marker validation pauper
Murringo	sweet	cultivar	*pauper*	marker validation pauper
Sulimo	sweet	cultivar	*pauper*	marker validation pauper
Neutra	sweet	cultivar	*pauper*	marker validation pauper
Nelly	sweet	cultivar	*pauper*	marker validation pauper
Peragis Stamm 749/40 × Frieda	bitter	F1 FiBL	*pauper* × *suavis*	check for F2 segregation analysis and breeding
Neuland × Nährquell	bitter	F1 FiBL	*exiguus* × *nutritius*	check for F2 segregation analysis and breeding
Neuland × Kievmutant	bitter	F1 FiBL	*exiguus* × *pauper*	check for F2 segregation analysis and breeding
Dieta × Frieda	bitter	F1 FiBL	*non-pauper* × *pauper*	check for F2 segregation analysis and breeding
Frieda × Peragis Stamm 749/40	bitter	F1 FiBL	*pauper* × *suavis*	check for F2 segregation analysis and breeding
Murringo × Butan	bitter	F1 FiBL	*pauper* × *exiguus*	check for F2 segregation analysis and breeding
Boros × Nährquell	bitter	F1 FiBL	*pauper* × *nutritius*	check for F2 segregation analysis and breeding
Breeding line1	sweet	breeding line FiBL	*pauper*	genotyping for breeding
Breeding line2	sweet	breeding line FiBL	*pauper*	genotyping for breeding
Breeding line3	sweet	breeding line FiBL	*pauper*	genotyping for breeding

Breeding line4	sweet	breeding line FiBL	*pauper*	genotyping for breeding
ex Murringo	sweet	breeding line FiBL	*pauper*	genotyping for breeding
Breeding line5	sweet	breeding line FiBL	*pauper*	genotyping for breeding
Breeding line6	sweet	breeding line FiBL	*pauper (?)*	genotyping for breeding

* to check homogeneity of these accessions, 1-2 additional plants of these accessions were included into the panel. Individual plants, where the Dragendorff alkaloid phenotype didn't match the expected one, were replaced or complemented by other individuals of the same accession but better matching phenotype.

Suppl. Table II - 2: Genotypes and phenotypes of a cross and the derived F1 and the F2 population, that segregates for two independent and recessive lQA genes, symbolised by A/a and B/b, with the capital letter representing the hQA allele respectively. The genotypes of the gametes of the F2 are given in paranthesis. The phenotypes of the genotypes are indicated by orange for hQA and light blue for lQA.

Generation		**Genotypes**				**Phenotypes**
Crossing formula:		aaBB × AAbb				lQA × lQA
F1:		AaBb				all hQA
F2:		(AB)	(Ab)	(aB)	(ab)	
	(AB)	AABB	AABb	AaBB	AaBb	7:9
	(Ab)	AABb	AAbb	AaBb	Aabb	lQA : hQA
	(aB)	AaBB	AaBb	aaBB	aaBb	
	(ab)	AaBb	Aabb	aaBb	aabb	



Suppl. Figure II - 1: Example of QA test with Dragendorff paper. The reaction was scored with 1 for a purely white stain, 2 for white stain with reddish encirceling, 3 for a pinkish stain with a lighter center and 4 for an intensely red stain without discoloration in the center.

Suppl. Table II - 3: Proximal sequences of SNP markers used for primer design. The first column indicates the chromosome and the location of the SNP. The column "ref/alt" indicates the reference allele and the alternative alleles of the SNPs.

SNP chr_ location	ref/alt	sequence of proximal region (100bp)	provenience	note
Lalb_Chr06_13713695	A/G	C[A/G]GTGACTGCCTTAACCGCATCACACCGCAATTGC[G/A]GTGTAACTCTGAATCACAATGATGTGGCCACATCAGGCCGCAACACAACCGTATCTGGCC[A/G]CAATGCGATCCCATTTTGATTGCATTTG[T/G]ACCGTATCGGACTGCATTGGA[C/G]CA[C/T]ATCAGGCTG[C/T]ATCGAATTCAAAAATATTGTTTTAATCACAAAATAATT	LfL_GWAS2	
Lalb_Chr10_16473189	C/T	TAAAGGAGTTGTAAATTTCGAAAAAACCCTAAAATAAAATATTAATTTATGCAGATATCAGAGAGGTCAAGGGGAAAGAGGAAGATGAAGATGGAGTTTTTGCCA[C/T]AGGGGTGAGACTCTTTTTTTAAATATTAAATTAATTATAAATTTAAATCTAAGTTATTCAATTTTTTATGACATGTATCATAATATTA[T/A]GTTT	LfL-GS	
Lalb_Chr12_2626435	C/T	CATCTTGATTCTTTAAAGAGTGAAATAACTGGACAACTGAAGGGGCTTTTTGGGTATTAGCTAGTTTTGCAGATGGCCGAGATGGAACTGGTGGTGGCAGCGG[C/T]GGAGGTAGTGGCAGGCTCGGCTTGGGTGGAGATTTCCTTCCTATACTTGAAGTTTCAGGAATGGCATGACTTGCTGCATGAGTTGGAATTGGTGT	LfL-GS	
Lalb_Chr14_5832122	G/X	GTTTTCTTGGGAAGGAAACAAAATTGCATATTATGTTTCTAATAAGAGTTTGTCTCAACATTATGGTACTCTATAAGCAGCTGATATTTCATTCCGGTTGGA[G/X]CCCCTCACTACTCTTTGGCTTCATAAGGTGAACAATCATGA[C/T]GTTGATGCTGC[T/A]TACCTAGATGCATAAGATAGTTTAAACTTTTGTTGGCGTGGA	LfL_GWAS2	
Lalb_Chr16_2668267	C/T	AAGCTGATGGAGTGCAGTGGGGACTCCAATGCTTTAGTTGATCCTGTTCAAATTGACTC[C/A]TGCGGCCCATATGATGTACAAGTTGAAAGCAGCATCAATAAAAGGGA[C/T]CGTGGAAATGTTGGTGTAGCCATTCAGCTAAATCAAGACATGGGTTTAGGTGCTTGTACTGAATCTGTATATGGAAATTTGGATACAGATACTGT	LfL_GWAS2	
Lalb_Chr16_12866499	T/N	TCTCATGTCTTTTGTGCCCTAGTAACTT[A/C]CATGTGTCCAGACTCCAGAATCTATTCTACTTTTCTCTCCACTTCCATGCTGCAACTAGTCCAATTTGTTGC[T/N]CTTCATTACCTCCTCACTATCTCCTGTTCTTTGGGAATGAACACTCTGAACAAC[T/G]TTTCCTTCGCTGCCACTACCACCTTTTCTATTATATGTCCGTG	LfL_GWAS2	partly (T/C)
Lalb_Chr18_12110444	A/G	AATTTATTTGTTTTTGGGTATTGTTTGTAGTGGGTTTATTATAGATTTAACATTTTTATGTTAATGGACCAGTTGTACTATCCCATTGCAAAGACTTTAGC[A/G]GAGAAAGCGGGCTGGGATTTTGCAAAGGAGACCGGTTTGGATGTAGTGATGATAAATCCTGGCACAGCCTTGGGCCCTCTCATTCCACCAAGAATCA	LfL-GS	
Lalb_Chr18_12359687	T/A	CTTTGAGAGGGCTTGTTTGAGCACAAAAGCAGGATCTAGTTGGCCATTTGAGGGTGATTGGTAAACATGAACAATAGGCCAAATGCTATCAGGATAGGGTCTATG[T/A]TCAAGTG[C/T]AGATAGAGGAAGAACAATGGAAGGAGTTGGTTTAGAAGGTTTAACATGCACTACTTCTTTCATCTCAATTTTCAGTGATGCCATTT	LfL_GWAS2 & LfL-GS	pauper
Lalb_Chr18_12457069	A/T	CCAAGATATGTCAGGGCTGCCTCCCAAAAGATCTCAGGACCATGCTATTCACTTAAAGGAAGGGGCTGGGATACCTAATCTCAGACCATATAGATGTCTCTAC[A/T]ACCAGAAAGCAGAGGTGGAAAGGTTAGTAAGGGAGATGCTGGAGGCTGGGGTTATTAAGCATAGTATCAGTCCTTACTCCAACCCCATTATACTGG	LfL_GWAS2 & LfL-GS	
Lalb_Chr18_12463263	A/G	ATTGGTAGCTTTATTAGCTAGCCG[G/A]TTTTTGAGTACCGACTTTTCGTAGAAATTTCTCTACGTATGATATTCCCATTGCCTCTTGGAATGAATAGGTACCC[A/G]GGATTACTGATCGTGGTTTGGTTCCAAGACCAGTGAAAAAGGTTGCCATCCTTGGTGGGGGACTAATGGGCTCCGGAATAGCAACTGCTTTAATTC	LfL_GWAS2	
Lalb_Chr21_12361245	A/T	ATTTGATGCAGCGTATGCACGATGGTTAGAGGAACATAATAGGCAAATCAATGAGTTAATG[A/G]CTGCAGTAAATTCTCACGCCGGAGATATTGAACTCCGTAC[A/T]ATTGTAGACAATGCCATTACACAATTTGATGAGATCTTCAGGCTAAAAAGCATTGCAGCTAAAGCCGATGTTTTCCATATTCTGTCTGGAATGTGG	LfL-GS	
Lalb_Chr21_12633426	C/A	AAGAAGACCAACTGGCACATACTTCAAAAGGTGGTATATATGGAAGGGACGAGAGCTCTGGCAGCAACATTAGGGAGGGAAGCAGCCCTGGGAACAAAGTTTCTGA[C/A]TCAGGATATTCCGTTCCGAAACATGCTAATTGGGATGGTGTTGAATTGACTGCTGTTTCAGAGATCAGT[T/A]GTAACCCACCTGATGATTCACG	LfL-GS	
Lalb_Chr21_12637843	T/C	CCGTGAAGTCACTGCTCTTGGCCATCTGTGCTGTTACACCCCCTTTTGCCACATAGATGGGTTGGCAACCACTGCGA[T/C]GCTGCTATCCCAGCGTTATAGTG[T/C]GCAATTGACTACTTAGGGCATGAGTTGACTAATATTACAAAAATGAACATCTTCCTCCTCACATAAGTTTGATT[T/C]TTTGATTCTTTTCTTGACAGGCTG	LfL-GS	
Lalb_Chr21_12637918	T/C	CTGCTATCCCAGCGTTATAGTG[T/C]GCAATTGACTACTTAGGGCATGAGTTGACTAATATTACAAAAATGAACATCTTCCTCCTCACATAAGTTTGATT[T/C]TTTGATTCTTTTCTTGACAGGCTGCCTCCTTCAACAAAAGGGTCAAGTAATGCTTTACATAAGCGCCCTCCAGTTTCTCGGGTTTTATCCTCCCCTGATAT	LfL-GS	
Lalb_Chr25_11617387	A/C	TAAATGGAATTTCTATGATTTACCCCTCGGCTATTTGTGGTGTCACTTTCCCTGTTTGATATTTATCTCTCTTCAGTGGTCTAGAGCTAATACATTAAATTCA[A/C]TGAAACAGAAAGTTCCTCGTATA[C/T]CTGATTAT[C/T]GAAATTGCAGCACCCATAGTCAGTGATAACTTTCTGGTAAAACTCAAGAATAACCAACACAAACTTG	LfL_GWAS2	
Lalb_Chr25_1199129	T/G	AATCAAATTTCTCTGTTAAATGGCATAATGGAACAATAAAAGTTACATAATTCATCTATTGAGTAAAAATGAATCCAGAACTAAAGATTTAGATTTGTCT[T/G]CATCACCTCTGTTGGAAAAGTATCAGAAGTATCGTGGTACAAACCAGATACCACATCTAAACTATCCCTTCCTTTCACTGTCTTCTCCAAGTTACTG	LfL-GS	

Suppl. Table II - 4: DNA extraction protocol, used for the bulk sequencing. (This protocol was given to me by Dr. Michael Schneider, who doesn't remember where he got it from.)

Material:	**Reagents:**
- 2 ml screw caps - 1.5 ml tubes - metal/ceramic beads - tissue lyser - heating block - table centrifuge - fume hood	▪ Extraction buffer ▪ Potassium Acetate 5M ▪ *RNAse (sequencing)* ▪ Isopropanol (icecold) ▪ Ethanol 70%

Procedure

1. Check buffer and sample preparation below. Have your samples in **2 ml screw caps**!
2. Preheat bottle with extraction buffer in a water bath to 65°C. Any precipitation should disappear by heating up. Do not pipette anything from the bottle into an epi or intermediate tube before the liquid is clear. Take subsample for your extraction into e.g. a falcon tube and preheat at 65°C on the heat block until you are ready to pipette the buffer to the samples.
3. Lysis: lyse dry leaves for 1 minute at full speed in tissue lyser. Precut leaf tissue in small pieces. Use 2 4mm grinding balls for the grinding step. See "Sample Prep" on Page 2.
4. Add **800 ul Extraction Buffer and** vortex until all tissue is evenly dispersed in the buffer.
5. Incubation on thermoblock: **65°C 10 min**
6. After incubation at 65 °C put thermoblock to 37°C. Add **5 ul of 1:5 diluted RNAse A**, shake good, incubate 2**0 min at 37°C. This step can also be performed as last step.**
7. Add 300 ul KAc and leave it for 15 min. Gently mix by tilting tube every few minutes.
8. Centrifuge for 10 minutes at 6.000 rpm.
9. Carefully pipette 600 ul of supernatant into new **tubes.**
10. Centrifuge for 10 minutes at 6.000 rpm.
11. Pipette 400 ul of supernatant into new **1.5 ml tubes** containing **400 ul ice cold Iso-propanol**. Mix gently by tilting for 2-3 min at room temperature. If successful or the DNA amount in the sample is high, the DNA should become visible by eye.
12. Centrifuge for 10 min at max speed. You now can see a white palette on the bottom of the tube. Pour out the iso-propanol into a waste container. The DNA will stick to the tube.
13. Add 600ul 70% ethanol and centrifuge for 10 mins at 1000 rpm.
14. Pour the ethanol into the waste container. Empty the waste container into funnel in fume hood.
15. Dry the DNA palette at room temperature (15 min) with open lids under the fume hood. For this place them upside down on a clean paper towel and lean them e.g. against a rack. The tubes should stand slightly tilted, so the ethanol can either flow out or directly evaporate. Check evaporation by smelling the tubes (ethanol smell indicates that the tubes must dry some more).
16. Re-suspend the DNA palette in 50-100 ul 1x TE buffer or water.

Preparation of Reagents:

Extraction Buffer:	**concentr.**	**Dilute**
25ml of Tris/HCl (pH 8.0) (Roth 5429.3)	1 M	3.03g /25 ml
31ml of NaCl (Roth 3957.1)	4 M	7.25g /31 ml
20ml of Na_2EDTA/HCl (pH 8.0) (Roth 8043.1)	0.5 M	3.72g / 20 ml
41ml of 10% SDS	10%	10g / 90 ml (not all is added)
Has to be pH 8.0. Mix all ingredients and make the volume up to 250ml by adding ELGA water.		

Potassium acetate 5M	**concentr.**	**Dilute**
KAc	5M	56.525g / 50ml

Elution buffer	**concentr.**	**Dilute**
Tris (Roth 5429.3)	10 mM	0.121 g / 100ml
(Don't add EDTA as it might interfere with PCR and sequencing)		

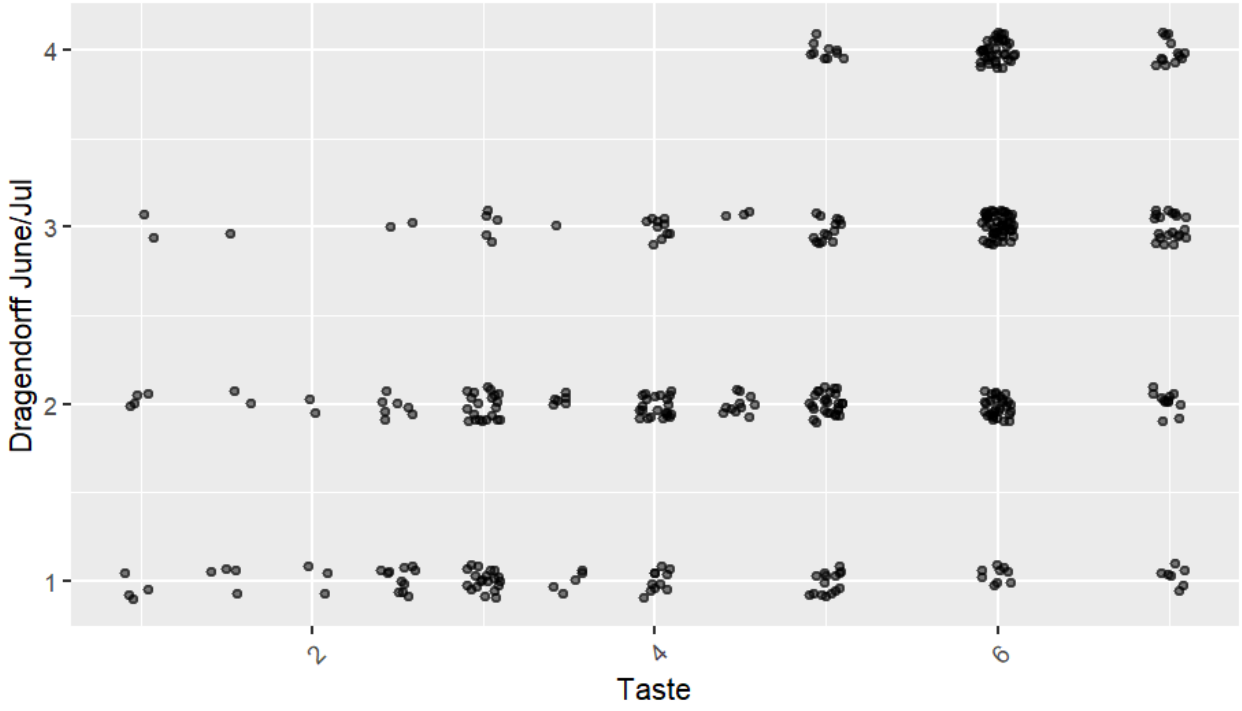
Suppl. Table II - 5: Dragendorff phenotyping and genotyping results of QA panel. Blue represents the reference allele, red the alternative allele and green the heterozygotes. The yellow highlight is the causal SNP for pauper, with the reference allele representing the "sweet" allele.

Accession	Dragendorff phenotype	genotype group	Lalb_Chr06_13713695	Lalb_Chr10_16473189	Lalb_Chr18_12110444	Lalb_Chr21_12633426	Lalb_Chr21_12637843	Lalb_Chr21_12637918	Lalb_Chr22_2016762	Lalb_Chr12_2626435	Lalb_Chr14_5832122	Lalb_Chr16_12866499	Lalb_Chr16_2668267	Lalb_Chr18_12457069	Lalb_Chr18_12463263	Lalb_Chr21_12361245	Lalb_Chr25_1199129	Lalb_Chr25_11617387	Lalb_Chr18_12359687
Gyulatanya	sweet	*pauper*	AA	CC	AA	CC	TT	TT	GT	CC	XX	TT	TT	TT	GG	AA	TT	AA	TT
Kalina			AA	CC	AA	CC	TT	TT	GG	CC	GG	TT	CC	AA	AA	AA	TT	AA	TT
elatus aus Neutra			AA	CC	AA	CC	TT	TC	GG	TT	XX	TT	CC	AA	AA	TT	TT	AA	TT
Breeding line 1			AA	CC	AA	AA	CC	CC	GG	CC	GX	TC	CC	AA	AA	TT	TG	AA	TT
Breeding line 4			AA	TT	AA	CC	TT	TC	GG	TT	XX	TT	CC	AA	AA	AA	TT	AA	TT
Feodora			AA	TT	AA	CC	TT	TT	GG	CC	GG	CC	CC	AA	AA	AA	GG	AA	TT
Murringo (original)			AA	TT	AA	AA	CC	CC	GT	TT	GG	TT	CC	AA	AA	TT	TT	AA	TT
breeding line 3			AA	TT	AA	CC	TT	TC	GG	CC	GG	CC	CC	AA	AA	AA	GG	AA	TT
Kiev mutant			AA	TT	AA	AA	CC	CC	GG	CC	GG	CC	CC	AA	AA	TT	GG	AA	TT
Kiev mutant			AA	TT	AA	AA	CC	CC	GG	CC	GG	CC	CC	AA	AA	TT	GG	AA	TT
breeding line 6			AA	CT	AG	CC	TT	TC	GG	CC	GG	TT	CC	AA	AA	AA	TT	AA	TT
Frieda			AA	CC	GG	AA	CC	CC	GG	CC	GG	CC	CC	AA	AA	TT	GG	AA	TT
Boros			AA	TT	AA	CC	TT	TT	GG	CC	GG	TT	CC	AA	AA	AA	TT	AA	TT
Amiga			AA	CC	AA	CC	TT	TC	GG	CC	GG	CC	CC	AA	AA	AA	GG	AA	TT
Sulimo			AA	CC	AA	CC	TT	TT	GG	CC	GG	TT	CC	AA	AA	TT	TT	AA	TT
exMurringo			AA	TT	AA	AA	CC	CC	GG	CC	GG	CC	CC	AA	AA	TT	GG	AA	TT
breeding line 5			AA	TT	AA	AA	CC	CC	GG	CC	XX	CC	CC	AA	AA	TT	GG	AA	TT
Peragis St. x Frieda	bitter	F1 plants	AA	CT	GG	CA	TC	TC	GG	CC	GX	CC	CC	AA	AA	TT	GG	AC	TA
Neuland x Nährquell			AA	CC	AG	CC	TT	TT	GG	CC	GX	TT	CC	AA	AA	AA	TT	AC	TA
Neuland xKievmut.			AA	CT	AG	AA	CC	CC	GG	CT	XX	TC	CT	AT	AG	AT	TG	AC	TA
Dieta x Frieda			AA	CT	GG	AA	CC	CC	GT	CC	GG	TC	CT	AT	AG	TT	TG	AC	TA
Frieda x Peragis St.			AA	CT	GG	CA	TC	TC	GG	CC	GX	CC	CC	AA	AA	TT	GG	AA	TA
Peragis Stamm 749	bitter	suppo sed sweet	AA	CT	AG	CA	TC	TC	GG	CC	GX	CC	CC	AA	AA	TT	GG	AC	TA
Butan			AA	TT	GG	CC	TT	TC	GG	CC	XX	TT	CC	AA	AA	AA	TT	AA	AA
Nährquell			GG	TT	GG	CC	TT	TC	GT	CC	XX	TT	TT	TT	GG	AA	TT	CC	AA
Nelly			AA	CC	GG	CC	TT	CC	GG	CC	XX	TT	CC	AA	AA	AA	TT	CC	AA
Neuland			AA	CT	GG	CC	TT	TC	GG	TT	GX	TT	CC	AA	AA	AA	TT	CC	AA
Neuland(bitter)			AA	CT	GG	CA	TC	TC	GG	CT	XX	TC	CT	AT	AG	AT	TG	CC	AA
LUP 2079	bitter	wildty pe	AA	TT	GG	AA	CC	CC	GG	TT	XX	TT	TT	TT	GG	AA	TT	CC	AA
Pop 2078			AA	TT	GG	AA	CC	CC	GT	TT	GG	CC	CC	AA	AA	TT	GG	CC	AA
Weibit			AA	TT	GG	AA	CC	CC	GG	CC	GG	CC	CC	AA	AA	TT	GG	CC	AA
Butan	sweet	non-paupe r	AA	TT	GG	CC	TT	TC	GG	CC	XX	TT	CC	AA	AA	AA	TT	CC	AA
Start			AA	CT	GG	CC	TT	TC	GG	CC	XX	TT	TT	TT	GG	AA	TT	CC	AA
Dieta			AA	TT	GG	AA	CC	CC	GT	CC	GG	CC	TT	TT	GG	TT	TT	CC	AA
Nährquell			AA	CC	GG	CC	TT	TC	GG	CC	XX	TT	CC	AA	AA	AA	TT	CC	AA
Nährquell			AA	CC	GG	CC	TT	TT	GG	CC	XX	TT	CC	AA	AA	AA	TT	CC	AA
Energy			AA	CT	AA	CC	TT	TT	GG	CC	XX	TT	CC	AA	AA	AA	TT	AC	TA
Neutra			GG	CC	AA	CC	TT	TT	GG	CC	XX	TT	TT	TT	GG	AA	TT	CC	AA
Zulika			AA	CC	AA	CA	TC	TC	GT	CC	GG	TT	CC	AA	AA	AT	TT	CC	AA
breeding line 2			GG	CC	GG	AA	CC	CC	GT	TT	XX	TT	NA	AA	AA	TT	GG	AA	AA
Murringo x Butan			AA	TT	AG	CA	TC	TC	GG	CC	GX	TC	CC	AA	AA	AT	TG	AC	TA
Boros x Nährquell			AA	CT	GG	CC	TT	TT	GG	CC	XX	TT	CC	AA	AA	AA	TT	CC	AA

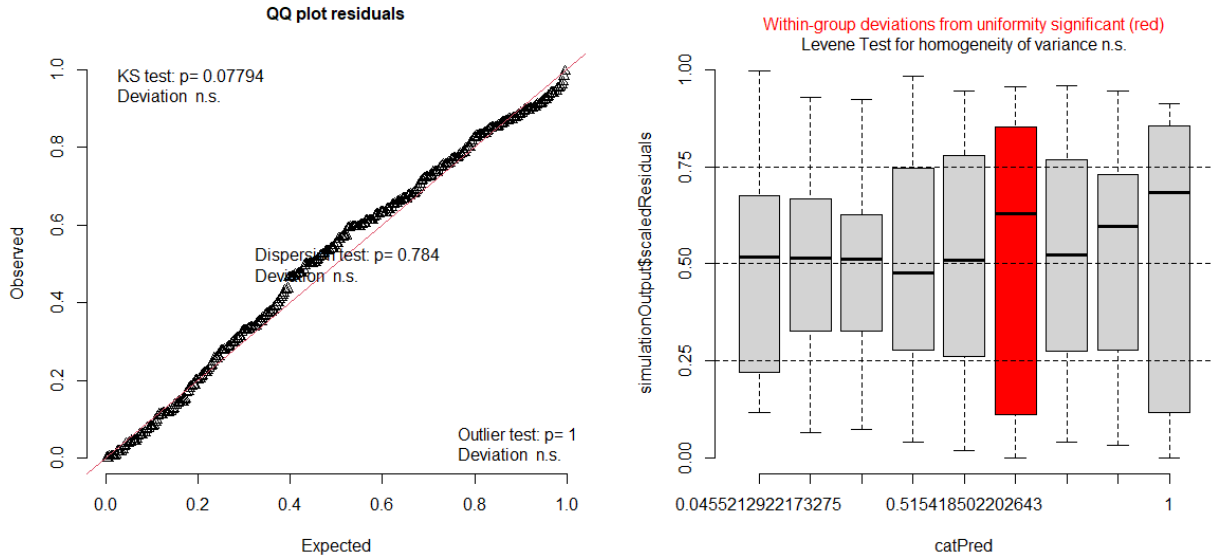
Suppl. Table II - 6: Genotyping results and combined QA scores of plants with the same genotypes as the parentals of the F2 populations. Blue represents the reference allele, red the alternative allele and green the heterozygotes. The yellow highlight is the causal SNP for pauper, with the reference allele representing the "sweet" allele.

Genotype	Lalb_Chr25_1161387	Lalb_Chr18_12359687	Lalb_Chr18_12463263	Lalb_Chr10_16473189	Lalb_Chr14_5832122	Lalb_Chr16_2668267	Lalb_Chr18_12457069	Lalb_Chr10_16473189	Lalb_Chr21_12637918	Lalb_Chr12_2626435	Combined QA score
Dieta	AA	AA	GG	CC		TT	TT				5
Dieta	AA	AA	GG	CC		TT	TT				6
Dieta	AA	AA	GG	CC		TT	TT				5
Dieta	AA	AA	GG	CC		TT	TT				5
Dieta	AA	AA	GG	CC		TT	TT				5
Dieta	AA	AA	GG	CC		TT	TT				5
Dieta	AA	AA	GG	CC		TT	AT				5
Dieta	AA	AA	GG	CC		TT	TT				5
Frieda*		AA	GG	CC		TT	TT				7
Nahrquell	AA	AA	AA		XX			CC	TT	CC	7
Nahrquell	AA	AA	AA		XX			CC	TT	CC	9
Nahrquell	AA	AA	AA		XX			CC	TT	CC	8
Nahrquell	AA	AA	AA		XX			CC	TT	CC	5
Nahrquell	AA	AA	AA		XX			CC	TT	CC	5
Nahrquell	AA	AA	AA		XX			TT	TT	CC	8
Nahrquell	AA	TA	GG		GG			CC	TT	CC	NA
Nahrquell	AA	TT	AA		NA			CC	TT	CC	NA
Nahrquell	AA	AA	AA		GG			CT	TT	CT	NA
Nahrquell	AA	AA	AG		GG						NA
Nahrquell	AA	AA	GG		GG						NA
Nahrquell	NA	AA	GG		XX						NA
Nahrquell	AA	AA	AA		XX						NA
Nahrquell	AA	AA	AA		XX						NA
Nahrquell	AA	AA	AA		XX						NA
Nahrquell	AA	AA	AA		XX						NA
Nahrquell	AA	AA	GG		XX						NA
Neuland	AA	AA	GG		XX			CC	CC	TT	11
Neuland	AA	AA	AA		XX			CC	TT	NA	NA
Neuland	AA	AA	AA		XX			CC	TT	NA	NA
Neuland	AA	AA	AA		XX			CC	TT	NA	NA
Neuland	AA	AA	AA		XX			CC	TT	NA	NA
Kiev_mutant	CC	TT	AA		GG						3
Kiev_mutant	CC	TT	AA		GG						4
Kiev_mutant	CC	TT	AA		GG						4
Kiev_mutant	CC	TT	AA		GG						5
Kiev_mutant	CC	TT	AA		GG						4
Kiev_mutant	CC	TT	AA		GG						6

*the QA score is high for a sweet cultivar and the genotype of this "Frieda" deviates from the Frieda genotype in the QA panel. This is likely a contaminant from the seed lot.



Suppl. Figure II - 2: Correlation of taste scores and Dragendorff test scores.



Suppl. Figure II - 3: Residual plots (package DHARMa) for the mixed model with Lalb_Chr18_12359687 and population as fixed and row as random variable: QA score ~ Lalb_Chr18_12359687 * population + row

Suppl. Table II - 7: ANOVA tables for marker candidates in different F2 populations.

F2 Neuland × Nährquell: Lalb_Chr14_5832122
Response: QA_score

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lalb_Chr18_12359687	2	414.40	207.200	23.9517	2.872e-09 ***
Lalb_Chr14_5832122	2	5.37	2.685	0.3104	0.7339
Residuals	103	891.03	8.651		

F2 Neuland × Kiev mutant: Lalb_ Chr18_12463263
Response: log10(QA_score)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lalb_Chr18_12359687	2	0.15407	0.077037	9.2003	0.0001926 ***
Lalb_Chr18_12463263	2	0.01368	0.006839	0.8168	0.4443151
Residuals	119	0.99643	0.008373		

F2 Neuland × Kiev mutant: Lalb_Chr25_11617387
Response: QA_score

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lalb_Chr18_12359687	2	109.00 54.501	13.7252	4.013e-06 ***	
Lalb_Chr25_11617387	2	2.42	1.212	0.3053	0.7375
Residuals	127	504.31	3.971		

F2 Dieta × Frieda: Lalb_Chr18_12457069
Response: QA_score

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lalb_Chr18_12359687	2	270.11	135.055	30.1988	4.362e-11 ***
Lalb_Chr18_12457069	2	17.30	8.652	1.9345	0.1496
Residuals	105	469.58	4.472		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Suppl. Table II - 8: Number of reads and sequencing depth for the bulked samples before and after data processing and filtering.

	raw data		**after quality filtering and mapping**		**after duplicate removal**	
Bulked sample	**number of reads (mio)**	**read depth**	**number of reads (mio)**	**read depth**	**number of reads (mio)**	**read depth**
Dieta	117.3	38.4	68.8	22.5	53.5	17.5
Frieda	113.4	37.1	68.2	22.3	54.4	17.8
sweet	97.1	31.7	59.7	19.5	46.8	15.3
bitter	99.1	32.4	60.4	19.7	48.0	15.7

Suppl. Table II - 9: Functional annotations for genes associated with haplotypes on Lalb_chr05 that differ between the sweet and the bitter bulk with log10(p-value) > 4.

start	Gene_ID	log10 pval	annotation (whitelupin.fr)
6382337	Lalb_Chr05g0222011	5.35	Putative transcription factor interactor and regulator LIM family
6691899	Lalb_Chr05g0222461	5.21	Putative transcription factor bHLH family
6587256	Lalb_Chr05g0222331	5.01	Putative cellulase
6599775	Lalb_Chr05g0222341	4.87	Putative pectinesterase
6609391	Lalb_Chr05g0222351	4.86	Putative phosphoglucomutase (alpha-D-glucose-1,6-bisphosphate-dependent)
6278389	Lalb_Chr05g0221801	4.73	Putative High mobility group protein HMGA
6638693	Lalb_Chr05g0222381	4.71	Putative aminoacyltransferase, E1 ubiquitin-activating enzyme
5799140	Lalb_Chr05g0221051	4.67	Putative transcription factor AP2-EREBP family
6731701	Lalb_Chr05g0222541	4.66	Putative protein NIM1-INTERACTING 2
6388537	Lalb_Chr05g0222021	4.55	hypothetical protein -BLAST- PREDICTED: IRK-interacting protein-like
6394782	Lalb_Chr05g0222031	4.41	Putative histidine kinase, Protein-serine/threonine phosphatase
6334316	Lalb_Chr05g0221891	4.39	Putative protein-synthesizing GTPase
6287319	Lalb_Chr05g0221811	4.36	Putative mitogen-activated protein kinase CMGC-MAPK family
6670560	Lalb_Chr05g0222421	4.33	Putative ubiquitinyl hydrolase 1
6665115	Lalb_Chr05g0222401	4.21	hypothetical protein
5926782	Lalb_Chr05g0221281	4.18	Putative pre-mRNA-splicing factor Cwf15/Cwc15
6509494	Lalb_Chr05g0222211	4.18	Putative nucleoporin, Nup155, WD40/YVTN repeat-like-containing domain-containing protein
7181929	Lalb_Chr05g0223181	4.17	Putative aminocyclopropanecarboxylate oxidase
6369081	Lalb_Chr05g0221971	4.15	Putative PB1 domain-containing protein
5999693	Lalb_Chr05g0221391	4.14	putative protein
7301404	Lalb_Chr05g0223301	4.10	hypothetical protein
6377416	Lalb_Chr05g0221991	4.03	hypothetical protein