Lupinus mutabilis:

exploiting diversity to guide breeding

Agata Gulisano

Propositions

1. *Lupinus mutabilis* is a competitive multipurpose crop for Europe.
   (this thesis)

2. Tailoring the re-domestication of minor crops for specific locations and needs, is key to a sustainable diversification of the agricultural system.
   (this thesis)

3. Academic research is the only field in which workers pay to sell their product.

4. Academic institutions should be obliged to offset their carbon emissions.

5. Large supermarket chains marked the end of our relationship with food.

6. It is easier to believe conspiracy than science.

Propositions belonging to thesis, entitled

'Lupinus mutabilis: exploiting diversity to guide breeding'

Agata Gulisano

Wageningen, 09 December 2022

Lupinus mutabilis: exploiting diversity to guide breeding

Agata Gulisano

Thesis committee

Promotors
Prof. Dr. Luisa M. Trindade
Personal Chair, Plant Breeding
Wageningen University & Research

Co-promotor

Dr. M. João Caldas Paulo
Researcher, Biometris
Wageningen University & Research

Other members

Prof. Dr. M.E. Schranz, Wageningen University and Research

Dr. S.C. Pyett, Wageningen University and Research

Dr. H. Muylle, Institute of Agricultural and Fisheries Research, Merelbeke, Belgium

Prof. Dr. B.J. Zwaan,Wageningen University & Research

This research was conducted under the auspices of the Graduate School of Experimental Plant Science (EPS).

Lupinus mutabilis: exploiting diversity to guide breeding

Agata Gulisano

Thesis

Submitted in fulfilment of the requirements for the joint degree of doctor at
Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr. A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 9 December 2022
at 1.30 pm in the Omnia Auditorium.

Agata Gulisano
Lupinus mutabilis: exploiting diversity to guide breeding, 180 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2022)
With references, with summary in English

ISBN 978-94-6447-464-0
DOI https://doi.org/10.18174/579563

Table of Contents

1

General Introduction

Grand challenges require major changes

Humankind is currently facing some of the most profound challenges it has ever faced: climate change, overpopulation, natural resource depletion, environmental pollution. As we witness dramatic increases in food and protein demands, but also in land degradation and biodiversity erosion, the need for a change of direction in our agricultural and food system has become more pressing than ever. The increased use of land, irrigation and agrochemicals that has accompanied the growth of agricultural production since the Green revolution, is now showing its drawbacks. The agricultural system we currently rely on, based on intensive monoculture cropping, is held responsible for almost a quarter of greenhouse gas emissions, represents the primary driver of biodiversity loss and accounts for approximately 70 % of all freshwater withdrawals globally, being responsible for huge areas of land and soil degradation (P. Smith *et al.*, 2013; Benton *et al.*, 2021; Food and Agriculture Organization, AQUASTAT data). Similarly, the high consumption of meat and dairy products characterizing our current diets, is responsible for a disproportionate share of environmental pressure regarding both resource utilization (land area, biodiversity and fresh water) and pollution (eutrophication, pesticides, climate change). Besides, livestock production relies to a great extent on the global trade of a single protein crop -soybean from tropical countries- with deleterious ecological and socio-economic effects on the long term (Boerema *et al.*, 2016). Overall, today's predominance of high-intensity systems focused on a relatively low number of commodity-crops, often outsourced, needs to be rethought if we are to reverse the current trends and ensure future prosperity to the growing population. A transition towards more sustainable diets, replacing meat-proteins with plant-based proteins, and the development of a sustainable and self-sufficient agricultural system appear pivotal. Novel production systems are needed that combine high productivity with increased crop diversity and enhanced resilience, aiming not only at reducing environmental harm but at actually improving ecosystem health (Brummer *et al.*, 2011).

In response to these challenges, the European Union (European Commission, 2012) and almost 50 countries around the globe (Fund *et al.*, 2018) have put forward the concept of bioeconomy, as a novel production system (Stegmann *et al.*, 2020). Bioeconomy has been defined as a system based on "*the production of renewable biological resources and the conversion of these resources and waste streams into value added products, such as food, feed, bio-based and bioenergy products"* (European Commission, 2012).These principles were implemented into a European Bioeconomy Strategy in 2012 and further updated in 2018, with the overall objective to have sustainability and circularity at the heart of bioeconomy (European Commission, 2018; Stegmann *et al.*, 2020). This strategy aims at ensuring food and nutrition security through a sustainable management of natural resources and a reduction on the dependence of non-renewable, unsustainable sources thereby contributing to mitigate climate change by replacing petrochemicals and fossil fuels products, and strengthening the competitiveness of Europe by creating new business opportunities and jobs (Figure 1). However, the focus should not only be on the use of biomass as building block for materials,

chemicals and energy to displace fossil resources, but also on fostering the protein transition and ensuring the sustainability of the production of biomass feedstock avoiding increased competition with food production over the use of arable land and water resources. Pivotal to the achievement of these goals appear the introduction of resilient crops, able to strive on marginal lands and improve it, and a tailored design of these new crops based on a biorefinery cascading approach.

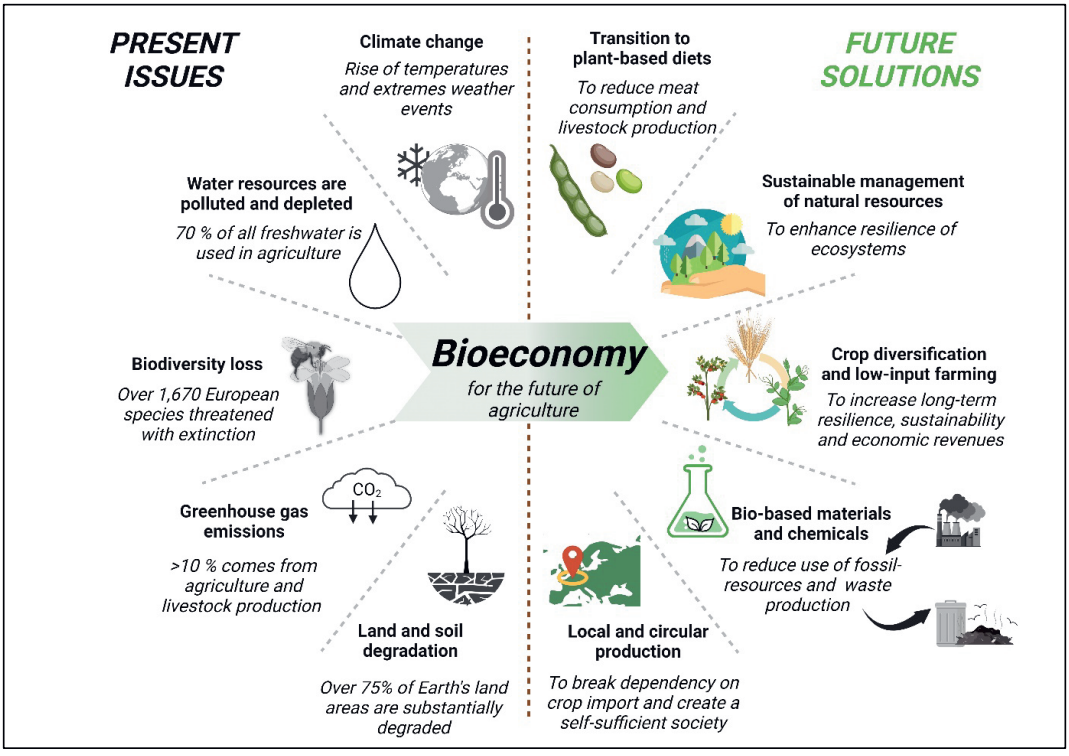


FIGURE 1| Present issues affecting the current agricultural system and future solutions offered by a Bioeconomy transition.

Producing more with less: the role of multipurpose crops and production residues in a sustainable bioeconomy

Within the framework of bioeconomy, one of the main strategies identified by Europe to improve the sustainability of our production system is the use of multipurpose crops and the valorisation of production residues (Schoenmakere *et al.*, 2018). Multipurpose crops, designed to ensure an optimal use of the different plant parts for food, feed, materials and energy generation, offer the clear advantage of producing more outputs from less inputs and that is the reason why

they should be designed to fit optimally a biorefinery cascading approach. Biorefining has been defined as the sustainable processing of biomass into a spectrum of marketable products and energy (S.R. Hughes *et al.*, 2013). A major task is to identify the most promising bio-based products to be co-produced, aiming at the co-production of different functional stream from one biomass source (e.g. protein, oil and an energy), to optimize overall process economics and minimize overall environmental impact (Odegard *et al.*, 2012; S.R. Hughes *et al.*, 2013). With the tools provided by molecular biology and chemical engineering, the types of co-products that can be derived from biomass may be almost limitless. Energy in the form of biofuel is one possible output, along with other bulk or commodity products of high volume and comparatively low value, such as industrial oils, adhesives, surfactants, solvents, and biopolymers for biodegradable fibres and plastics, in addition to low-volume, high-value products, such as chemicals for the food, cosmetics, and pharmaceutical industries (Bastos Lima, 2018).

In addition, within agriculture considerable amounts of biomass are still currently discarded as residues of the production process, instead of being transformed into further products. More than half the globally harvested biomass consists of agricultural residues and inedible biomass, such as cereal and legume straw, vegetable crop stalks, leaves and shoots or tree pruning and litter. It is estimated that 121 million tonnes of agricultural crop residues (mainly straw) could be generated annually in Europe, together with 46 million tonnes of forestry residues and 31 million tonnes of grass (Y. Iqbal *et al.*, 2016). The use of these residues as a resource for materials could create economic value by extending the biomass resource base. However, when increasing the economic use of crop residues, appropriate measure to safeguard the ecosystem quality should be taken into account as these residues can contribute substantially to the stability of biomass production when integrated in the soil by providing protection against erosion, increasing the capacity to store water, recycling nutrients and providing organic matter (Smil, 1999).

Unfortunately, current crops have been bred and designed in time with different scopes in mind and are now far from optimal for multipurpose uses. Traditional breeding for high-yielding varieties has so far mainly focused on maximizing the yield of a specific product from each crop, optimizing them to use a single fraction of the total biomass. Instead, we now require the development of crops in which biomass and agricultural residues have been designed/bred to be economically viable to process into bio-products. The focus should be on the precursor carbohydrates, lignin, oils, and proteins, and the combination of biotechnological and chemical conversion processes to transform these substances into valuable products (Birgit Kamm *et al.*, 2007). The use of secondary metabolites is also of enormous interest for the industrial and scientific sector, as these phytochemicals can find high-value applications in pharmaceutical industries, cosmetics, nutrition, but also in manufacturing of drugs, dyes, fragrances, and in agriculture as antibacterial and bio-stimulant agents (Guerriero *et al.*, 2018). Main challenges to the exploitation of these substances include overcoming biomass recalcitrance, providing fair prices for crops and agricultural residues, and tailoring crops and production to specific environments (S.R. Hughes *et al.*, 2013).

Plan(t)s for the future of Europe

Breaking the dependency on soybean import

As discussed above, the needs to transition towards a more sustainable agricultural system and towards a plant-protein based diet, are urgent and concomitant. Hence, the EU and several EU Member States as individual countries have launched and adopted bioeconomy policy strategies to achieve long-term sustainable development and foster the protein transition. In order to build such a resilient system, one of the main focuses is on finding alternative to soybean, as Europe has become strongly dependent on the import of this crop from other countries. Currently, Europe produces only 29% of the high-protein commodities it needs, while 87 % of the remaining deficit is met by imported soybean and soymeal (Watson *et al.*, 2017). Every year Europe import approximately 30 million tonnes of soybean products, in the form of raw soybeans, soymeal and soybean oil mainly from Brazil, the United States, and Argentina. The access of European farmers to these source of low-cost soybean and soymeal has allowed an exponential increase in livestock production in the past sixty years, causing a drastic increase in greenhouse gas (GHG) emissions (Watson *et al.*, 2017; Tallentire *et al.*, 2018). The consequent increase of soybean production in South America, going from 0.26 to almost 60 million hectares between 1961 to 2020, has also led to deforestation and ecosystem conversion of the Amazon tropical rainforest, causing major habitat losses for plant and animal endangered species and releasing over 200 million tonnes of GHG (Voora *et al.*, 2020).

Re-introducing grain legumes as a sustainable alternative

Whilst growing soybeans in Europe is non-competitive with imports due to relatively low yields and a long growing season (Van Krimpen *et al.*, 2013) but also to higher production costs, the reintroduction of higher share of grain legumes* in the European cropping system has been amply suggested over the last decade. Grain legume production in Europe has witnessed a steep decrease in the past 70 years, mostly driven by rising competitions with cheaper imports and by the substitution of legume intakes by meat products (European Commission, 2018). The focus on cereals since the Green revolution, has also contributed to the neglect of legumes with the development of N-responsive varieties, deepening the dependence on chemical fertilizers and overshadowing the contribution of legumes and their well-known synergy with cereals (Ferreira *et al.*, 2021). It is now widely acknowledged that grain legumes have the potential to improve the agronomic performance of cropping systems by providing protein-rich food and feed and contributing to reduce dependence on imported protein (Watson *et al.*, 2017). Furthermore, grain legume production can significantly contribute to ecosystem services due to its low reliance on synthetic fertilizers, decreased GHG and increased diversification of crop rotation, leading to

*The term grain legume is used to indicate plants belonging to the Fabaceae family which are grown primarily for their edible seeds.

increases in above and below-ground biodiversity and changes in weed, pest, and disease pressures, as well as improvements in soil fertility and carbon storage.

The reintroduction of grain legumes into the European food system currently demands public financial and academic/technologic support, as well as changes in consumers dietary habits (Ferreira *et al.*, 2021). More research considering legumes cultivation methods and techniques (e.g., genetic trait selection) is needed, together with improvements in farmer's knowledge (Ferreira *et al.*, 2021). During the European Horizon 2020 program multiple dedicated plant-based protein projects were funded, subsidizing research on different grain legume species, including peas, fava beans, lupins, chickpeas, lentils, and soybeans. Some of these projects, have a specific focus on the introduction of new protein crops where agricultural side streams can be valorised into added-value products for the European biorefinery industry. The LIBBIO Horizon2020 research project is an example of such an integrated approach. It focuses on lupin and aims to develop and optimize a breeding and cropping program for Andean lupin *(Lupinus mutabilis*) on European marginal land, to develop consumer food, feed, non-food and bioenergy products (LIBBIO, 2017). The choice of this crop was driven by the competitive quality of its grains, with a protein and oil content similar to soybean, by the ability of this legume to fix nitrogen, mobilize soil phosphate and having low nutritional requirements for cultivation, and by its adaptation to different and adverse climatic conditions.

This thesis, as part of the LIBBIO project, investigate natural variation in an ample collection of *L. mutabilis* accessions to develop a set of tools for breeding new genotypes adapted to European conditions, with a focus on developing multipurpose varieties where agricultural side streams can be used in the biorefinery industry in an economically viable manner.

Lupinus mutabilis, an Andean crop with high potential

Lupinus mutabilis L. is a diploid [2n=48], mainly self-pollinating, legume species from the Andes. It has been cultivated, processed and consumed for more than 1500 years in Ecuador, Peru, Bolivia and parts of Colombia and Chile, where it still represents an important indigenous food crop mostly known as Andean lupin, chocho or tarwi (F. Carvajal-Larenas, 2013). The Andean Lupin has an excellent potential as a new alternative crop because of its protein (44 % dry weight) and oil content (18 % dry weight), which exceeds that of any other lupin species but also of crops such as oil seed rape, sunflower and soybean (Pate *et al.*, 1985; Blanco-Galdos, 1982). Moreover, *L. mutabilis* has also an enormous nutritional and nutraceutical potential in comparison to other legumes. The protein fraction is mainly composed by globulins (α-,β-,γ- and δ- conglutin, 91-94%), known for their cardiovascular health benefits and used in the control of insulin resistance and diabetes (Magni *et al.*, 2004; Belski *et al.*, 2011). In the fatty acid profile, linolenic and linoleic acid prevail, which also play an important role in human metabolism, contribute to lower blood pressure and to reduce the risk of cardiovascular disease and tumour (Naghshi *et al.*, 2021; Froyen *et al.*, 2020). In addition, lupin seeds are practically devoid of starch, gluten free and rich in

calcium, iron, zinc and dietary fibres. However, they are characterized by the presence of toxic alkaloids which confer a bitter taste and can lead to poisoning. Hence, alkaloids need to be properly removed through a "debittering process" prior to ingestion, limiting direct consumption for both human and animals. Nonetheless, these alkaloids can be also highly valuable, both in the medical field (they have anti-arrhythmic and anti-inflammatory effects among others), as bio-stimulant in agriculture, or as anti-bacterial and biocidal agents replacing synthetic toxins (Bunsupa *et al.*, 2012; Przybylak *et al.*, 2005; Romeo *et al.*, 2018; Bermúdez-Torres *et al.*, 2009)

In the past, the strong adaptation of this plant to a wide range of conditions favoured its expansion along the west side of South America, from Colombia to Northern Argentina, encompassing a wide range of latitudes and altitudes. However, the introduction of western legumes after the Spanish conquest brought big changes in dietary habits and a drastic reduction in the cultivation of *L. mutabilis* which got restricted to small farmer's field, mostly for their own needs (Caligari *et al.*, 2000). Nowadays *L. mutabilis* remains of agricultural importance only in Ecuador, Peru and Bolivia and has only recently gained the attention of Australia and Europe because of its high potential as a protein/oil crop.

Lupinus mutabilis belongs to the genus *Lupinus* L. which comprises around 267 species, grouped in two geographically isolated groups: 11 'Old World' species from Mediterranean and East African regions, and ~250 'New World' species from America (Drummond *et al.*, 2012). In the Andes, lupin species are particularly diverse and reach among the highest speciation rates known for any plant species (C. Hughes *et al.*, 2006). *L. mutabilis* was domesticated 2600 years ago in the highlands of north Peru and is currently the only American member of the genus Lupinus that has been domesticated and cultivated (Blanco-Galdos, 1982). Cultivated *L. mutabilis* plants are characterized by an indeterminate growth habit, a highly branched architecture, large seeds, multi-coloured flower and an annual life cycle (Clements *et al.*, 2008). They grow at altitudes between 2000 and 3850 m, in sandy and low fertile soil with a slightly acidic pH values comprised between 5 and 7. Water requirements are limited for this crop (350-850 mm) and it behaves as a photoperiod-independent plant, even though it is mainly cultivated in short day conditions. The adaptability of this species to the different microclimates and harsh conditions of the Andean region including extreme daily variations in temperature, strong winds, intense ultraviolet radiation, strong rains, hail, snow, high altitudes and low-fertile soils has significantly contributed to the persistence and importance of *L. mutabilis* as a subsistence crop for the local populations (Jacobsen *et al.*, 2008; Tapia, 2015). Furthermore, due to its nutritional qualities, *L. mutabilis* has represented an excellent and cheaper substitute to animal proteins while being also an important component of highland crop rotations in the Andes, contributing to soil fertility via nitrogen fixation and phosphorus mobilization (Lambers *et al.*, 2013).

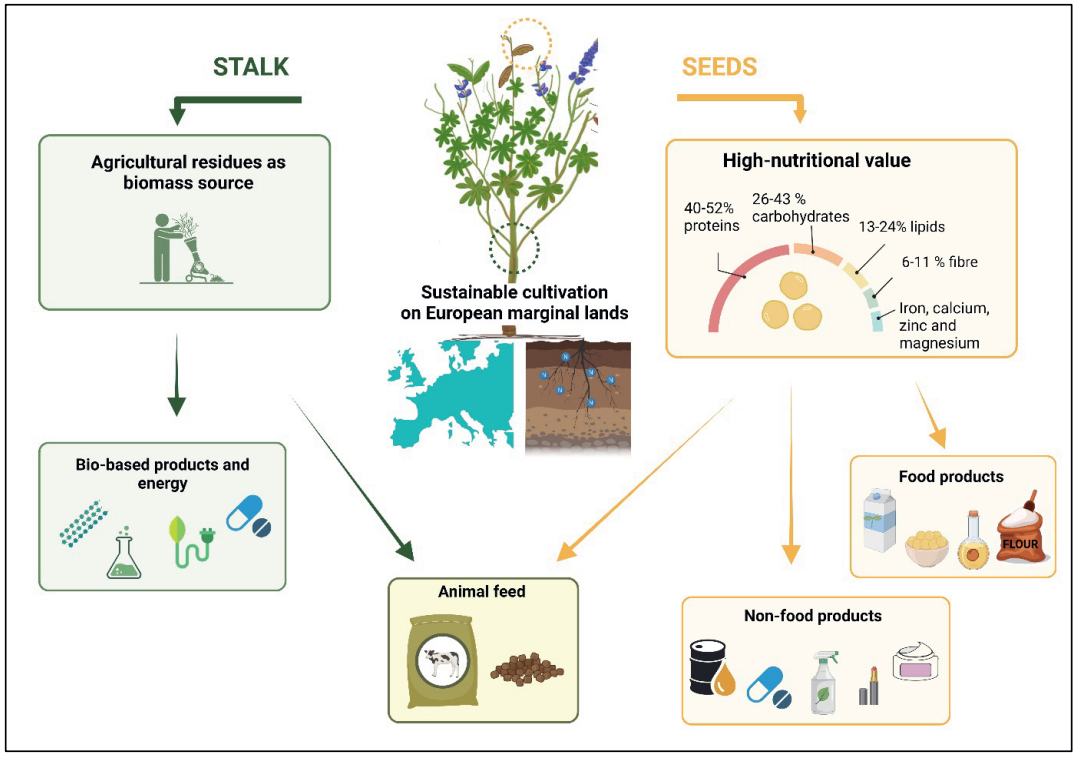


FIGURE 2 | A schematic representation of *L. mutabilis* use as a multipurpose crop.

The recent domestication of *L. mutabilis*, followed by a drastic reduction in cultivation and a limited economic importance on the global market, have led to a lack of research and breeding in this crop which has solely relied on selection activities by Andean farmers for the past 1500 years. Research on *L. mutabilis* has been carried out only in the last 50 years in Peru and other Andean countries, focusing mainly on the development of sweet varieties without alkaloids and on the selection of local ecotypes (Gulisano *et al.*, 2019). In Europe, studies to evaluate the potential of cultivating *L. mutabilis* started approximately at the same time, once the potential of the crop became well known. In these studies, they have mostly focused on the use of induced mutations and intraspecific mutant crossing, resulting in the use of limited sets of lines with limited genetic diversity. Preliminary evaluations of the performance of *L. mutabilis* in Europe pointed out the need of breeding for a better plant architecture and for early maturity, in order to increase seed yield (Caligari *et al.*, 2000). Years of mutation experiments in Poland have currently resulted in research lines characterized by improved yield, sweet seeds and determinate architecture while in Australia, male sterile lines have been used to introduce early vigour and resistance to pests (R Galek, 2010; R. Galek *et al.*, 2017; Sweetingham *et al.*, 2006). Still, the lack of extensive research in plant breeding and of advanced biotechnological methods in genetics, molecular cytogenetics or tissue cultures has limited the possibility of exploring natural variability and performing distant crosses and haploidization of material of reproduction. A coupled use of

germplasm and modern approaches to broaden the genetic base could now aid the introgression of desirable adaptive traits for specific environments, which are essential to develop *L. mutabilis* into a valuable crop outside the Andes.

Introducing *Lupinus mutabilis* as a multipurpose crop for Europe: outline of this thesis

This thesis aims to explore natural variation of *L. mutabilis* using for the first time a large collection of accessions including both germplasm material from the Andean region and breeding lines from European institutes. We investigate the adaptation of this collection in two contrasting environments in Europe: as a summer crop in North-Central European conditions and as a winter crop in Mediterranean conditions, and compare it with the adaptation to native conditions in the Andean region. By evaluating large germplasm collections in contrasting environments representing major growing conditions, we aim to identify useful genetic resources, adaptation strategies and tools to breed *L. mutabilis* varieties for specific environments. We used phenotyping and biochemical approaches to characterize and provide novel insights into the variability of this crop and performed genome-wide association studies to develop a set of tools for the breeding of new genotypes adapted to specific European conditions. Having as final aim the introduction of *Lupinus mutabilis* as a multipurpose crop on European marginal land, we focused not solely on agronomic traits and seed quality (protein/oil content) but expanded our investigation to the quality of the agricultural residues produced and evaluate the potential of using them as feedstock for the biorefinery industry. Characteristics such as intermediate growth, early maturity, high seed yield, high protein/oil content, and high quality of lignocellulosic biomass represent the main breeding targets to ensure the introduction of *L. mutabilis* as a multipurpose crop in Europe.

To address these objectives, this thesis has been structured as follows:

Chapter 2 reviews the benefits of advancing *L. mutabilis* as a crop for Europe and identifies current limitations and new breeding challenges for its development as a multipurpose crop. Origin and distribution, as well as biological and genetic features of this species are here described. Furthermore, an overview of past breeding achievements is presented and future prospects regarding the use of new molecular tools and available germplasm resources for breeding are discussed.

In **Chapter 3**, we explore the natural genetic variability of morphological and phenological traits of 225 *L. mutabilis* accessions. The collection is evaluated both in the native environment (Andean region) and over two cropping conditions in Europe, paving the way for the selection of accessions adapted to specific environments. Valuable insights into the relationships between key traits as flowering time, yield and architecture-related traits under different cropping conditions

are provided. This chapter also evaluates the potential of this selected panel for mapping studies and provides novel insights into the usefulness of specific germplasm resources from the Andes.

In **Chapter 4**, we highlight the potential of using *L. mutabilis* agricultural residues as a lignocellulosic feedstock for Europe. The quality of *L. mutabilis* lignocellulosic biomass and the underlining genetic structure are here investigated for the first time. Biomass quality of 223 accessions growing in contrasting cropping conditions in Europe was evaluated based on the biochemical analysis of fibre fractions and monosaccharide components of cell wall polysaccharides. Then, a high-throughput sequencing of the panel and the development of a genome-wide association study (GWAS) were used to identify quantitate trait loci (QTL) and candidate genes for biomass quality, that can lead to the development of markers for a common use across locations and of markers for specific locations.

A similar approach was used in **Chapter 5**, to identify QTL and candidate genes associated with relevant agronomic traits, namely flowering time, yield, and architecture-related traits. The genome-wide association study, performed over the two cropping conditions in Europe and the native environment, serves also as a starting point for deeper investigations of the mechanisms regulating the intricate system interconnecting plant development and phenology in this species.

In **Chapter 6**, we characterized seed quality of our *L. mutabilis* collection in the most suitable environments for seed yield, i.e. Mediterranean-winter conditions. This is the first characterization of total protein and oil content and oil composition assessing such a large share of germplasm material from the Andean region, providing vital information for the selection of valuable breeding material.

Finally, in **Chapter 7**, we discuss how the knowledge generated in this thesis can contribute to the development of *L. mutabilis* varieties adapted to different cropping conditions in Europe and, overall, to the development of *L. mutabilis* as a multipurpose crop. In this regard, we discuss how our findings add to the current body of knowledge about *L. mutabilis*, critically assess the limitations of our study and offer insightful recommendations to guarantee the success of future breeding programs, seeking to promote Andean lupin in the context of sustainable bioeconomy in Europe.

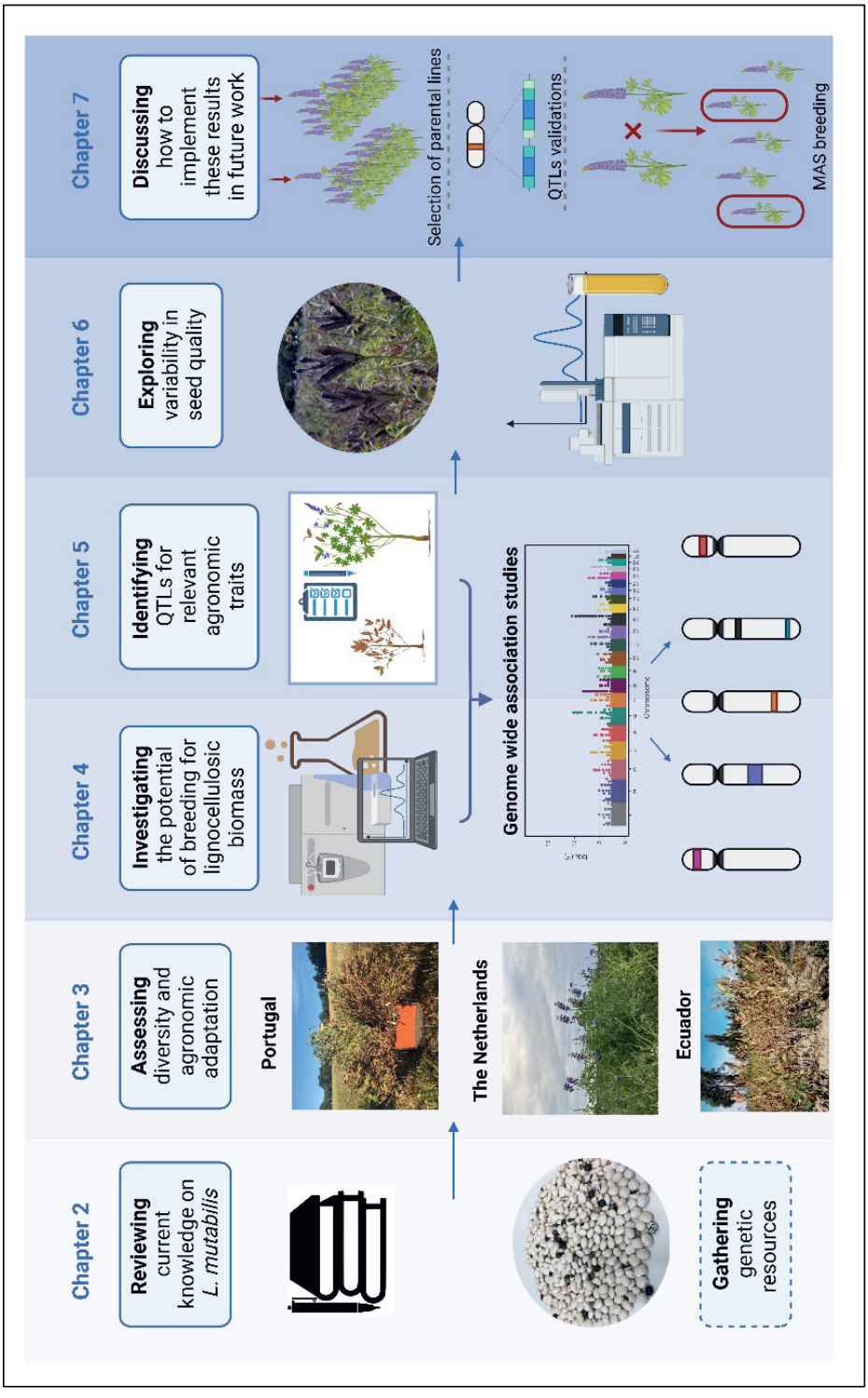


FIGURE 3| Graphical summary of the chapters and goals of this thesis.

2

Genetics and Breeding of *Lupinus mutabilis*: An Emerging Protein Crop

Agata Gulisano [1], Sofia Alves [2], João Neves Martins [2] and Luisa M. Trindade [1]

[1] Wageningen University & Research Plant Breeding, Wageningen University, Wageningen, Netherlands,
[2] DRAT, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, Lisbon, Portugal

Published as: Gulisano A, Alves S, Martins JN and Trindade LM (2019). "Genetics and Breeding of Lupinus mutabilis: An Emerging Protein Crop", **Frontiers in Plant Science,** doi: 10.3389/fpls.2019.01385

Abstract

Protein crops have gained increasing interest in recent years, as a transition towards plant-protein based diets appears pivotal to ensure global food security and preserve the environment. The Andean species *Lupinus mutabilis* emerges as an ideal protein crop with great potential for Europe and other regions with temperate climates. This species is characterized by oil and protein content similar to soybean and is highly valued for its adaptability to colder climates and low input agriculture on marginal land. However, its introduction outside the Andes has yet to take off. To date, *L. mutabilis* remains an under-studied crop, lacking high yield, early maturity and a consistent breeding history. This review paper identifies *L. mutabilis* limitations and potential uses, and suggests the main breeding targets for further improvement of this crop. It also highlights the potential of new molecular tools and available germplasm resources that can now be used to establish *L. mutabilis* as a viable protein crop.

Keywords: *Lupinus mutabilis*, lupin, breeding, genetics, protein crop, plant protein

Introduction

Over the past decades, challenges such as food security and environmental sustainability have earned the status of main priorities worldwide and are the basis of the 17 Sustainable Development Goals (SDGs) defined by the United Nations in 2015. As world population continues to rise, our food production has already exceeded the planet's environmental boundaries driving climate changes, biodiversity loss and unsustainable use of land and water. The growing demand for animal proteins has played an important role in this process, by turning livestock sector in the main user of agricultural land and in one of the biggest contributors to climate change. In light of the current situation, a transition from meat-intensive diets towards plant proteins-based diets is vital to ensure global food security and preserve the environment.
To create alternatives to animal protein, the cultivation of protein crops has gained interest in recent years. The European Union has launched initiatives to reduce its dependency on the import of soybean by growing an increasing quantity and variety of protein crops within the European member states. Research has focused on identifying sources of proteins that can reduce the current protein deficit while contributing to the transition to more sustainable agricultural systems. One such source is protein-rich leguminous plants. Legumes also stand out for their great potential in the reclamation of poor and marginal lands for agriculture, due to their ability to fix nitrogen and their beneficial effects on the soil (De Ron *et al.*, 2017). Among legumes, lupins have been identified as particularly promising, characterized by high-quality protein content, suitability for sustainable production and potential health benefits (Lucas *et al.*, 2015). The genus Lupinus includes almost 300 species, but only four play an important role in agriculture: *L. albus*, *L. angustifolius*, *L. luteus* and *L. mutabilis* (Gresta *et al.*, 2017). The first three listed species are native to Europe and represent the majority of lupins cultivated worldwide. At the same time, despite years of extensive research, the success of these species has been hampered by unstable yields, low oil content and adaptation to a narrow-range of agro-climatic conditions. The fourth species listed, *L. mutabilis*, is a species native to the Andes, and is cultivated only in some parts of South America and not yet commercially available in Europe (Lucas *et al.*, 2015). However, *L. mutabilis* is characterized by the highest grain quality of all cultivated lupins, with an oil content similar to soybean, and is adapted to low input farming in temperate climates. The combination of these characteristics makes *L. mutabilis* a potentially superior alternative to the current plant-based sources of protein and oil in Europe and other regions with temperate climates.

Lupinus mutabilis Sweet is considered to be one of the lost crops of the Incas. Its seeds are characterized by a high protein and oil content (44 % dw and 18 % dw, respectively), which exceeds that of any other lupin species (Blanco-Galdos, 1982; Pate *et al.*, 1985). In addition, lupin seeds are practically devoid of starch, and the major carbohydrates found are oligosaccharides (mainly stachyose and raffinose) and cell wall storage polysaccharides (Trugo *et al.*, 2003). Most essential amino acids, lysine in particular, are also present in the seeds (Table 1) together with a

substantial amount of dietary fibre and fatty acids (Table 2) (F.E. Carvajal-Larenas *et al.*, 2016). The history of this species as a subsistence crop in the Andes demonstrates its potential as a crop for low input agriculture on marginal lands. *L. mutabilis* shows a high adaptability to temperate and cold climates, low-fertile soils, high altitudes and harsh conditions while actively enriching the soil with nitrogen (Cowling *et al.*, 1998). Currently, its cultivation is mostly confined to the Andean region of South America, where its bitter seeds represent a regionally important food known as *tarwi*. It is an economically accessible source of good quality protein, on par with animal proteins, to a large percentage of the population (F. Carvajal-Larenas, 2013).

TABLE 1| A comparison of the essential amino acids profiles (+ cystine) of four species of lupins and soybean (*Glycine Max*)

	L. mutabilis	**L. angustifolius**	**L. albus**	**L. luteus**	**Glycine Max**
Histidine	3.5	2.6	2.0	3.1	3.8
Isoleucine	4.2	4	4.1	3.6	n.a.
Leucine	7.0	6.9	6.8	7.8	7.2
Lysine	5.8	4.6	4.5	4.5	5.4
Methionine	0.8	0.7	0.7	0.6	1.2
Phenylalanine	3.5	3.7	3.4	3.7	4.9
Threonine	3.5	3.4	3.4	3	5.4
Tryptophan	0.8	0.9	0.9	0.9	n.a.
Valine	3.8	3.7	3.8	3.4	4.9
Cystine	1.6	1.6	1.5	2.4	1.5

Data are expressed as g/100g of proteins (Carvajal-Larenas et al., 2016; Prakash and Misra,1988). (n.a.= data not available)

The presence of toxic alkaloids in the seeds and low yields (800-1300 kg/ha) have strongly limited the expansion of this crop (Tapia, 2015). Selection activities by Andean farmers in the past 1500 years of cultivation have represented the only means of domestication for *L. mutabilis*, leading to semi-domesticated forms characterized by non-shattering pods, large seeds, multi-coloured flowers, highly branched architecture and a more or less annual life cycle (Clements *et al.*, 2008).

TABLE 2| Nutritional composition of four species of Lupinus as compared to Soybean (*Glycine max*).

	Crude protein	Crude lipids	Crude Fibre	FA saturated/ unsaturated	Unsaturated Fatty acids (g/100g DW)			
					C18:1 (Oleic)	C18:2 (Linoleic)	C18:3 (Linolenic)	C22:1 (Erucic)
L. mutabilis	43.3	18.9	8.2	0.17	46.4	33.1	2.5	-
L. albus	38.2	11.2	8.9	0.5	54.0	18.7	8.6	0.4-2.7
L. luteus	42.2	5.5	15.8	0.13	28.5	48.2	6.3	*tr*-1.5
L. angustifolius	33.9	6.3	16	0.23	33.9	40.3	5.6	0.1-0.5
Glycine max	42.9	19.8	5.1	0.18	22.8	50.8	5.9-8.3	-

Data are expressed in g/100 g DW (F.E. Carvajal-Larenas *et al.*, 2016; Collins *et al.*, 1957; Hudson *et al.*, 1983; Sharma *et al.*, 2014; Prakash *et al.*, 1988).
* Crude fibre: insoluble residue, primarily composed of cellulose and lignin
(*tr* = in traces, less than 0.1 %)

It played an important role as a rotation crop in Andean agriculture, but the introduction of western pulses during the Spanish conquest in the sixteenth century, led to its decline and marginalization (Caligari *et al.*, 2000; Cowling *et al.*, 1998). In contrast, the wide genetic diversity that characterizes this crop has enabled its adaptation to poor soils and microhabitats, preserving its cultivation in many areas where other crops cannot grow (F. Carvajal-Larenas, 2013). This genetic diversity is also reflected in a broad phenotypic diversity, e.g. of seeds and flowers colour (as shown in Fig.1).

In recent years, efforts have been made to re-establish *L. mutabilis* as a crop in South America, and to also adapt it to conditions in Europe (Caligari *et al.*, 2000). Numerous studies investigating the nutritional profile and potential applications of these grains have found a wide range of possible products ranging from proteins, oil, and food additives to cosmetics, medicines and bio-pesticides. In contrast, few studies have addressed the agronomic aspect of *L. mutabilis* cultivation. From these studies it emerges that the main obstacle to *L. mutabilis* cultivation is the lack of high yielding, early maturing genotypes. These results are mainly determined by an indeterminate growth habit and a lack of locally adapted genotypes, and can be overcome via breeding (Caligari *et al.*, 2000). To date *L. mutabilis* remains an under-studied crop, characterized by a very young and fragmented breeding history. The important role that this crop could play in the transition toward a more sustainable food production system has prompted us to review the current state of this crop. This paper summarizes past breeding achievements and sheds light on the new breeding challenges we must resolve to establish *L. mutabilis* as a protein crop in Europe.

Origin and distribution of the "Andean lupin"

The earliest archaeological evidence of domesticated *L. mutabilis* seeds has been found in Mantaro Valley, central Peru and dates back to ca. 1800 BP. The use of RADseq in the analysis of this archaeological material confirms that *L. mutabilis* was first domesticated not far from the Montaro Valley in the Cajamarca region (north Peru), from the wild progenitor *L. piurensis*. Demographic analysis suggest that *L. mutabilis* split from its progenitor around 2600 BP (650 BC) and suffered a domestication bottleneck and a subsequent rapid population expansion as it became cultivated across the Andes (Atchison *et al.*, 2016). *L. mutabilis* presence has been reported across the eastern side of South America, from Colombia to the North of Argentina (from 10°N to 20°S), and over a wide range of altitudes, from 1500 to 3800 m a.s.l. (Jacobsen and Mujica, 2008). The crop is adapted to a temperate climate and is strongly influenced by day length. It is susceptible to low temperatures (-2 °C) in the initial stages, and requires about 350-800 mm of rainfall and can grow for 240-300 days (Adomas *et al.*, 2015; Jacobsen *et al.*, 2006; Jacobsen and Mujica, 2008). Based on these requirements, *L. mutabilis* could be cultivated in Southern Europe as a winter crop, and in Northern Europe as a summer crop. Nowadays *L. mutabilis* is of agricultural importance only in Ecuador, Peru and Bolivia. Approximately 1895 ha are cultivated in Bolivia with an average yield of 648 kg/ha, 5974 ha in Ecuador (400 kg/ha) and 10628 ha in Peru (1335 kg/ha) (Mercado *et al.*, 2018).

FIGURE 1| Phenotypic variation in flowers and seeds of *L. mutabilis*.

Biological and genetic features

L. mutabilis is an annual herbaceous plant of the Fabaceae family. It is an autogamous species, with hermaphroditic flowers arranged in apical racemes, but characterized by a predominant level of allogamy. Different ranges of cross-fertilization by insects have been reported, fluctuating

from 4-11 % in Peru to 9.5-18.9 % in Poland (Blanco-Galdos, 1982; Gnatowska *et al.*, 2000). It has been observed that multiple groups of insects visit *L. mutabilis*, suggesting that this species could be a generalist; bees of Apidae family and bumblebees from the genus Xylocopa are the main visitors in native environments, while bumblebees from the genus Bombus are more common in Europe (Ochoa-Zavala *et al.*; Arnold *et al.*, 2014). The isolation of different genotypes is thus indispensable in breeding programs, as much as the careful wrapping of emasculated flowers in intraspecific hybridization (Adomas *et al.*, 2015; E Von Baer, 2011).

Phylogenetic analysis places *L. mutabilis* (2n=48) within the Andean clade of Western New World species of the genus Lupinus. This genus includes almost 300 species, grouped by their different centres of origin into Old World (Mediterranean) and New World (American) subgenera. To date, *L. mutabilis* is the only cultivated species from the New World group (Gresta *et al.*, 2017). Notably, the Andean clade to which it belongs is characterized by the highest speciation rate within the genus (C. Hughes and Eastwood, 2006). The species in this clade belong to a paleoploid group of plants with basic chromosome number x=6 (Naganowska *et al.*, 2003). Events of allo- and autopolyploidization, together with other chromosomal rearrangements, during the evolution of this species might have led to duplication/ or triplication of genome regions, as observed in the Old World species *Lupinus angustifolius* (Kroc *et al.*, 2014).

The unexploited potential of *L. mutabilis*, an under-studied crop

L. mutabilis appears to be a valid alternative to soybeans for satisfying plant protein requirements in Europe. Like soybean, *L. mutabilis* seeds are rich in proteins as well as in oil. They can find applications as food and feed, but also as raw materials for the production of bio-based products. On the other hand, *L. mutabilis* cultivation tolerates better cold climates and can therefore contribute to the production and diversification of sustainable european sources of proteins and oil.

However, despite the clear potential, research on *L. mutabilis* has been limited. As is often the case for under-utilized crops, *L. mutabilis* has been long neglected by research and industry due to its limited economic importance on the global market. A recent domestication and a breeding history fragmented in time and space have also contributed to this neglect, resulting in a lack of genetic improvement and inferior yield. In the Andes, *L. mutabilis* germplasm collection and breeding programs started only in the 1970s and have so far relied on participatory approaches with farmers for the selection of local ecotypes (Table 3). The selection of genotypes with better yields mainly relies on the geographical distribution and vegetative cycle of the ecotypes and it has rarely resulted in the registration of cultivars (Fries *et al.*, 2007; Peralta *et al.*, 2012; Vicente Rojas, 2016). In Europe, researchers began working on the selection of sweet lines in the 1920s, but it was only in the 1970s, when the nutritional value of *L. mutabilis* seeds became well known, that the interest for this crop arose. The difficult accessibility of germplasm from the Andean area was overcome in Europe with a large use of induced mutations and intraspecific crossing of mutants. Preliminary field trials of *L. mutabilis* in Europe reported large differences in seed yields, from 0.5 to 6.5 t/ha depending on years and location (Masefield, 1976; Romer *et al.*, 1992; Rubenschuh, 1997; Weissmann *et al.*, 1992). In 1993, the first European project aimed at evaluating the "Adaptation of *Lupinus mutabilis* to European soil and climate conditions" was funded. Field trials reported very low seed yields (1.1 t/ha) and pointed out the need of breeding for a better plant architecture and early maturity (Caligari *et al.*, 2000). Many years of mutation experiments in Poland have resulted in improvement in yield and sweetness and in the selection

of determinate lines for research purposes, but not yet in the establishment or registration of new varieties (R Galek, 2010; R. Galek *et al.*, 2017) (Table 3). Australia has also shown interest in *L. mutabilis* and multiple projects to evaluate its potential for southern Australia were funded. Of particular relevance in their work was the selection of male sterile lines, used to introduce early vigour, anthracnose resistance, tolerance to brown spot and resistance to cucumber mosaic virus in *L. mutabilis* (Sweetingham *et al.*, 2006) (Table 3). The ongoing development of recombinant inbred lines (RIL) population at the University of Western Australia is mentioned in the literature and could be exploited for mapping QTL, however little information is available about its existence and state (J. C. Clements and M. N. Nelson, unpubl. data in (Berger *et al.*, 2013).

In Europe, the urgent need to provide alternative protein sources and recover marginal land has contributed to revive the interest in *L. mutabilis*. Recently, a new program investigating *L. mutabilis* cropping in marginal lands for enhanced bio economy has been funded under the European Union's Horizon 2020 program (www.libbio.net). The possibility of cultivating *L. mutabilis* as a summer crop in North-central Europe and winter crop in the Mediterranean area is being investigated, along with the development of pre-industrial processing and the assessment of its socio-economic and environmental impact.

Establishing *L. mutabilis* as a protein crop in Europe: the breeding challenges

Adaptation to European Environment

The environmental differences between the native environment of *L. mutabilis* and other cultivation areas around the world, such as Europe, represent one of the barriers to the expansion of this crop. In temperate climatic conditions *L. mutabilis* cultivation is characterized by a long period of maturation and uneven maturation of the pod, blossom drop, and shattering of early stage pods (R Galek *et al.*, 2007; A. Hardy *et al.*, 1998; Swiecicki *et al.*, 2004). Due to low resistance to frost during the first growth stage, sowing is limited to autumn in Mediterranean environments and to spring in northern countries. In both cases, the crop will reach flowering towards the beginning of the dry season. Dry conditions can accelerate maturation, but considerably affect the biomass yield, pod set and consequently, the seed yield (A Hardy *et al.*, 1997). Hence, it is crucial to generate early maturing genotypes with increased drought tolerance and consistent yield performance (yield stability). Previous work has pointed out that vernalization has no effect on early and mid-season genotypes of *L. mutabilis*, but can reduce the flowering time of late season genotypes such as Inti. It was observed that a vernalization period of 2-4 weeks at 6 °C can shorten flowering time of four weeks in Inti, reducing the gap between early- and late-flowering lines to only 3 weeks (Kedar N. Adhikari *et al.*, 2012).

Research on drought stress in *L. mutabilis* has uncovered the existence of different drought tolerance strategies across genotypes, either via stomatal adjustments or through the accumulation of osmoprotectants. Some traits, like stomatal conductance and water potential, appear to decrease uniformly among all accessions while other traits such as membrane ion leakage or accumulation of proline and soluble sugars show particular trends depending on the genotype. This might indicate the ability of some *L. mutabilis* genotypes to adapt their cell membrane during periods of water stress, as an alternative strategy to stomatal adjustment (Lizarazo *et al.*, 2010). Therefore, both stomatal conductance and membrane ion leakage can prove useful in the selection of drought resistant cultivars.

TABLE 3 | A list of *L. mutabilis* lines involved in breeding research.

Area of selection	Line	Characteristics	References
Chile	Inti*	Stable cultivar with 0.0075 % alkaloid content in seeds, but low yield and long vegetation period.	(Gross *et al.*, 1988) (E Von Baer, 2011)
Bolivia	Chumpi, TarwiNawi	Ecotypes grown in Potosi, characterized by dark brown seeds.	(Vicente Rojas, 2016)
	Tolarapa, Dulce	Ecotypes grown in the area of Cochabamba.	(Gross *et al.*, 1981)
	Carabuco*	Variety inscribed in the National Register of seeds. Characterized by early maturing and white seeds with a cuboid flat shape.	(Vicente Rojas, 2016)
Ecuador	I-450 Andino* I-451 Guaranguito*	Early maturing genotypes (6 months), uniform white seeds and higher yield (1370 kg/ha on average). Susceptible to anthracnose. Registered by INIAP.	(Peralta *et al.*, 2012)
	ECU-2700 ECU-2658	Genotypes selected for resistance to anthracnose and high yield (1445 kg/ha on average).	(Guaytarilla *et al.*)
Poland	KW-1	Completely determinate mutant, with no lateral branches. Characterized by tall growth, liability to lodge and low seed production.	(Römer, 1994)
	Research lines	Genotypes with shorter growth period, reduced number of branches and lower alkaloid content obtained combining intraspecific crosses with induced mutation.	(Sawicka, 1993) (Stawinski *et al.*, 2001) (Galek, 2010)
Australia	ID13, ID18, ID32,ID33, JC243,P28725	Advanced low alkaloid, breeding lines to assess adaptation of the species to eastern states.	(Clements *et al.*, 2008)
	P27033	Male sterile line	(Sweetingham *et al.*, 2006)
	P25954	Restorer line	(Sweetingham *et al.*, 2006)
	P26961	Early line	(Kedar N. Adhikari *et al.*, 2012)
	P27808	Mid-season line	(Kedar N. Adhikari *et al.*, 2012)
Russia	KVIR2381	Russian breeding line used in crosses to introduce tolerance to brown spot and resistance to cucumber mosaic virus (CMV).	(Sweetingham *et al.*, 2006)

**L. mutabilis* lines that developed into cultivars.

Response to photoperiod is another important factor for determining adaptation to different locations. Reports on photoperiodic sensitivity in *L. mutabilis* are contrasting. Hackbarth reported *L. mutabilis* as neutral to day length, while Jacobsen and Mujica affirm that in the Andean region *L. mutabilis* accelerates grain filling when the day length is short (Hackbarth, 1961; Jacobsen and Mujica, 2008).Given the latter, adaptation at higher latitudes should be based on the selection of lines less sensitive to day length effects on grain filling. A more exhaustive understanding of sensitivity to day length through characterization of germplasm collections and knowledge about

its genetic basis would enable the generation of genotypes for high latitudes with little or no sensitivity (Jacobsen and Mujica, 2008).

Growth habit: toward a semi-determinate type

Indeterminate growth habit and sympodial branching pattern have been identified as the main factors limiting yield of *L. mutabilis* in European field trials (Caligari *et al.*, 2000). In *L. mutabilis* the vegetative development begins with the production of a main stem bearing a terminal inflorescence and continues with the production of successive orders of branches throughout the entire growing season (that can be from 0 to 52 branches), as long as growing conditions are favourable (Blanco-Galdos, 1982; A Hardy *et al.*, 1997) (Fig. 2A). This growth habit leads to an overlap of vegetative and reproductive phases, characterized in this species by a preferential partitioning of nutrients to vegetative growth. As a result, the possibility of uniform maturation is hindered and reproductive growth is constantly delayed, often so far as to coincide with late-season drought, thus further reducing productivity. Furthermore, only the racemes of the main stem and first order branches are highly productive, while the production becomes progressively weaker on the other order of branches (K.N. Adhikari *et al.*, 2001).

It has long been proposed that the development of determinate lines could guarantee a more stable seed yield by providing an earlier and more uniform maturation (Huyghe, 1998). Determinate lupin cultivars have been obtained in *L. albus, L. luteus and L. angustifolius* mainly through selection of spontaneous or induced mutants. Vavilov's homologous order of restricted branching (*rb*) was selected independently in these species with a different mode of inheritance, a different number of alleles in the *rb* locus and somewhat differentiated expression in the respective species (Gorynowicz *et al.*, 2014). Similarly, a completely determinate line of *L. mutabilis* was found upon induced mutation with EMS. This mono-stem determinate mutant - *L. mutabilis* KW 1- did not produce lateral branches, matured early and was characterized by tall stems (Sawicka, 1993) (Fig. 2C). The inheritance of the determinate character was found to be monogenic recessive (Römer, 1994). Unfortunately, determinate plants were found to lodge and were not able to compensate for stress during main stem flowering because of increased pod set on the branches (Romer, 1995). Semi-determinate types with only one or two orders of lateral branches up to the top of the plant thus seem preferable from an agronomic point of view (Caligari *et al.*, 2000) (Fig. 2B). Ongoing research in Poland has focused on crossing the KW-1 mutant with early maturing mutants characterized by a reduced number of side branches for the selection of determinate form for research purposes (R Galek *et al.*, 2007; E. Sawicka-Sienkiewicz *et al.*, 2006; E.J. Sawicka-Sienkiewicz *et al.*, 2001). Determinate forms distinguished by medium-tall stems without lateral branches, resistant to lodgings and with early generative growth have been obtained (R Galek *et al.*, 2007). Indeterminate forms appear to have a higher mass of stems and plant aerial parts and a lower share of seeds in the yield structure, and may therefore be more suitable for biomass production (Adomas *et al.*, 2015; Gas, 2014).

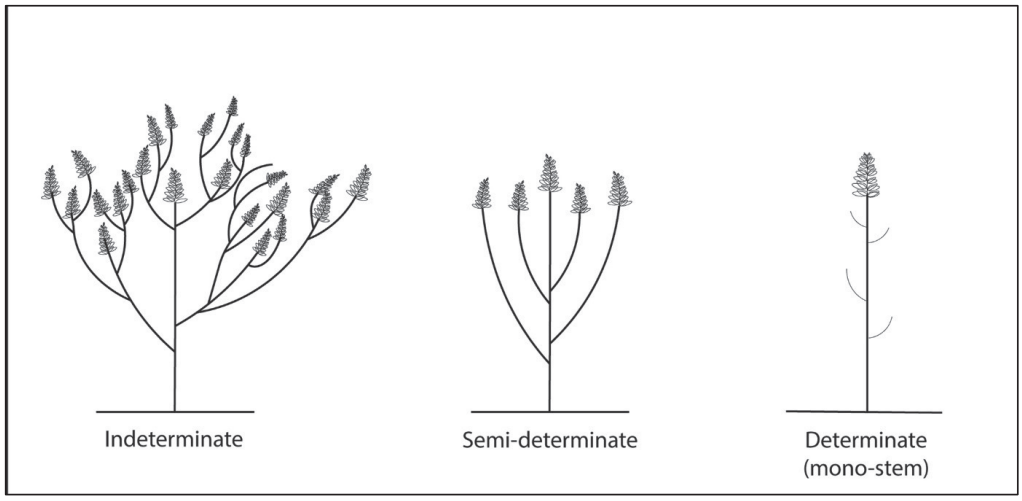


FIGURE 2| Different growth forms of *L. mutabilis*.

Understanding the mechanisms regulating alkaloid content

Food and feed industries have set the strict threshold of 0.02% (DM weight) alkaloid content in lupin seeds (Cowling *et al.*, 1998; Frick *et al.*, 2017). Quinolizidine Alkaloids (QAs) are typically synthesized by lupin species and are mainly known for causing bitter taste and anticholinergic toxicity when present in the grains. However, QAs also play an important role in the mechanism of defence against pathogens and predators, have allelopathic functions (Michael Wink, 1993) and constitute nitrogen reserves for the plant (M Wink *et al.*, 1985). They are biosynthesized from L-lysine in green tissues of the plant, transported via phloem and stored in all the organs of the plant, especially seeds. The content and composition of QAs depend on many factors, including genotype, biotic/abiotic stresses and pedoclimatic conditions. Each lupin species is characterized by a different alkaloid profile, known as an alkaloid- fingerprint, which fluctuates among the different organ of the plant, expressing a lower diversity and concentration in leaves than in seeds (Boschin *et al.*, 2013; Michael Wink *et al.*, 1995). Although the chemistry of Quinolizidine alkaloids has been extensively studied leading to the identification of more than 170 structures (Michael Wink, 1993), their biosynthetic pathway is only partially elucidated and information on the genes and enzymes involved remains limited (Frick *et al.*, 2017) .

The breeding of sweet lines of *L. mutabilis* has been mainly based on the selection of natural and induced mutants, mostly in Chile, Poland and Australia. The first stable "sweet" variety, Inti, was bred in Chile in 1980. It was characterized by an alkaloid content of 0.0075% with no reported detrimental effect on the protein (51%) or oil (16%) content, but low yield and long vegetation period hindered its adoption in different places (Gross *et al.*, 1988; E Von Baer, 2011). Yet, the inheritance of the trait was recessive and of polygenic nature, such that only 12% of the F2 plants had low alkaloid content (E. Von Baer *et al.*). These characteristics require major efforts to maintain the purity of mother's lines and to prevent the risk of progressive re-bittering due to cross-pollination in regions where lupin grows in the wild (Santana *et al.*, 2001; E Von Baer, 2011). In 1984, seed treatments with ethyl methanesulfonate led to the identification of the recessive allele *mutal* of the gene *Mutal*. When homozygous, the allele *mutal* was found to reduce the alkaloid level to 0.2-0.3% of seeds DM, giving rise to plants organoleptically sweet both in their seeds and vegetative parts (Williams *et al.*, 1984). It has been suggested that along the reselection process

additional minor alleles were recombined at several loci to lower alkaloid levels (Clements *et al.*, 2008). At present, none of the mutations found has led to complete suppression of alkaloids. The reduction in total alkaloids is mainly due to a reduced percentage of sparteine and lupanine, the two most toxic QAs to humans (Williams *et al.*, 1984) . The result of the work done in Chile in the last 40 years is the acquisition of a new variety, PINTA (Inti x SCG9) which combines low content of alkaloids, high content of protein and oil, and large seeds (E Von Baer, 2011). In Poland, post-mutagen treated material has been widely screened using iodine test to select 13 lines that don't exceed 0.1 % of alkaloid content in the seeds. These genotypes can be very useful in breeding programs, particularly to derive homozygous lines. In addition to facilitating the development of stables sweet varieties these homozygous lines can also be used to study the inheritance of alkaloid content in seeds (R. Galek *et al.*, 2017).

A major drawback of reducing alkaloids is the increased sensitivity of plants to pests and diseases. Future work should therefore targets the development of bitter/sweet lines, with sufficient level of alkaloids in the vegetative tissues to deter pathogens, but low levels in the seeds (M. Wink, 1990). To use this strategy fundamental knowledge on how to target the transporters involved in the translocation of QAs from source tissues to seeds is required. Candidate transporters may include plasma membrane importers in cells of reproductive tissue, and vacuolar membrane importers in cells of both aerial and reproductive tissues, as alkaloids are often sequestered within vacuoles to avoid toxic effects within tissues (Yazaki *et al.*, 2008). To our knowledge, there are no studies yet investigating these mechanisms.

Seed color, a matter of acceptance

To further develop the market for *L. mutabilis*, it is essential to take into account consumer preferences. When whole lupin beans are marketed as food, seed coat colour becomes a decisive trait for the acceptance of a cultivar. As for *L. mutabilis*, white colour is the most attractive for consumers. The phenotypical diversity in seed shape and seed coat colour observed in this species appear to be larger than that in all the other lupins (Blanco-Galdos, 1982). Seed characteristics with large diversity include shape (from lenticulate to spherical), primary seed colour, secondary seed colour and its pattern distribution (Fig.1). The colour can vary from pearly white to solid black, and include beige/yellow, brown, dark brown and intermediate colors, like brownish green and greyish colours. Most seeds have a secondary colour distribution in darker tones of the primary colour. The secondary colour distribution also varies between a large range of patterns, such as moustache, eyebrow, crescent, marbled, or spotted which can be expressed singularly or in combination (Falconí, 2012; Tapia, 2015). The variability in seed coat colour may reflect the genetic pressure *L. mutabilis* was subjected to during its domestication, but very little is known about the genetic mechanism behind this trait. Some authors try to explain this variation hypothesizing the concerted effect of different alleles in the control of different colours and at different regions of the seed coat. That is, having different genes controlling the primary colour, the secondary colour, the colour of the hilum and/or its adjacent region and the different patterns of distribution of secondary colours (Blanco-Galdos, 1982). Another possible explanation for the existence of such diversity in seed colour and patterns may be the presence of transposable elements, as observed in other crops (X. Li *et al.*, 2012). There appears to be a connection between seed colour and flower colour. Darker seeds lead to darker flowers, suggesting that the white colour behaves as a recessive character (Blanco-Galdos, 1982).

The complexity in seed colour represents a great challenge for breeders to select pure lines with uniform and heritable colours and patterns, and in particular to combine locus for high yield and

white colour. Still, pearly white is in 95% of the cases the most common colour found in the cultivars sampled for germplasm collections of the Andean regions.

Identification of health-promoting proteins

Lupinus mutabilis seeds contain a high content of protein, ranging from 38 to 45 % of DM; yet, the identification of unique properties in *L. mutabilis* proteins opens the door to new markets and raises the nutritional and economic value of the crop. The major protein classes encountered in legume seeds are globulins and albumins, followed by minor fractions of prolamin and glutelin (Doxastakis, 2000). Globulins (α-,β-,γ- and δ- conglutins) represent about 91-94 % of the proteins in *L. mutabilis,* while albumins only ~6.4 % (Santos *et al.*, 1997). Interest in conglutins has exponentially increased since their beneficial nutritional and pharmaceutical properties have been shown, such as cardiovascular health benefits and the use of γ- and β-conglutin in the control of insulin resistance and diabetes as well as anti-inflammatory molecules (Belski *et al.*, 2011; Lima-Cabello *et al.*, 2017; Magni *et al.*, 2004). (Foley *et al.*, 2015) used 16 individual conglutin genes previously identified in *L. angustifolius* to characterize homologous genes in five other lupin species, including *L. mutabilis*. Oddly, transcriptomic studies revealed the lowest level of conglutin transcripts for *L. mutabilis,* but the highest percentage of proteins. The expression levels for β-conglutin were particularly high (~40 %) and for γ-conglutin exceptionally low (4%), while the expression levels of α- and δ- conglutin (~26% and 30%) were comparable to the values encountered in *L. albus* and *L. angustifolius*. Previous studies have highlighted considerable differences in structure and composition of α-conglutin and β-conglutin in *L. mutabilis* as compared to *L. albus*. In the case of α-conglutin differences were observed also within different genotypes of *L. mutabilis* (Inti and Potosi) (Santos *et al.*, 1997), suggesting that these proteins may have different functions between and within lupin species (F.E. Carvajal-Larenas *et al.*, 2016). In contrast , γ-conglutin was reported to possess identical composition in all lupin species studied and to represent approximately 6 % of the total proteins in *L. mutabilis* seeds (F.E. Carvajal-Larenas *et al.*, 2016). Regarding albumins in *L. mutabilis*, they were found to be less abundant and different in structure when compared to *L. albus* (Santos *et al.*, 1997). Finally, the presence of ferritin (Fe-rich protein) in the protein profile of lupin (Strozycki *et al.*, 2007) increases the nutritional value of this crop by offering a safe way to increase dietary iron intake. The success of its use in the development of food products for special nutritional purposes would depend on the achievement of ferritin overexpression, which may result in easier, cheaper and more accepted methods for increasing dietary iron intake than supplementing and/or fortifying other crops (Zielińska-Dawidziak, 2015).

Exploiting the high nutritional value of *L. mutabilis* oil

L. mutabilis seeds are also an important source of oil. The oil content of this species (~18%) is the highest within lupins and the only one comparable to soybean (20%). Moreover, its fatty acid composition is nutritionally superior to that of soybean: both have a similar ratio of saturated/unsaturated fatty acids (17-18%), but *L. mutabilis* has a lower amount of linolenic acid, thus avoiding the need for industrial removal of this acid as soybean and *L. albus* do, and its oil stability is naturally higher (Schoeneberger *et al.*, 1982). In addition *L. mutabilis* oil does not have any toxic erucic acid found in other lupin species, and when compared to other edible oils presents a higher or similar quality, being inferior only to olive oil (Martins *et al.*, 2016) (Table 2). Improvement of oil production via breeding could further enhance the economic suitability of this crop by making it dual-purpose for protein and oil, in a manner similar to the soybean (Lucas *et al.*, 2015).

Oil content and composition are influenced by both genetic and environmental factors, and previous studies have identified a large environmental component. One study by (Williams, 1979) has reported higher oil content in late-flowering and late–maturing varieties and identified a highly significant correlation between oil content and the length of interval between flowering and pod maturity. Negative correlations between protein and oil content are also reported in the literature ($r = -0.71$; $r = -0.77$) (Clements *et al.*, 2008; Jacobsen and Mujica, 2006; Perez *et al.*, 1984).

The identification of accessions in which oil and protein content are not (or less) inversely related could make it possible to combine high levels of both components in the seeds through selective breeding (Romer *et al.*, 1986). An opportunity could come from the fibre component of lupin seeds, mainly β-galactan chains in the form of thickened cell walls of the endosperm (Al-Kaisey *et al.*, 1992). Since catabolism of both carbohydrates and lipids generally represents the main source of germination energy, it is possible to assume that oil content might be increased via breeding at the expense of β-galactan content.

Relevant resources for future breeding of *L. mutabilis*

Germplasm collections to exploit natural diversity

The Andean region, centre of origin and domestication of *L. mutabilis,* represents the main hotspot of diversity for this species. Germplasm collections were started in 1974 by Dr. Oscar Blanco at the University of Cusco (Peru) and soon extended to Bolivia and Ecuador. At present, South American institutions hold more than 3000 genotypes of Andean Lupin. The largest and most relevant germplasm collections of *L. mutabilis* are held in the gene banks of Peru, Ecuador and Bolivia, but smaller collections are also present in Chile, Argentina, Colombia, Australia, Russia, Poland, Germany, Spain, Hungary, United Kingdom and Portugal. Yet, reports suggest much of the diversity remains uncollected (Jacobsen and Mujica, 2008). The presence of a considerable variation across germplasm is shown by different phenotypic traits, such as a wide range of growing periods, branching patterns, colour and shape of grains and flowers, and flowering times. Both Inter Simple Sequence Repeats (ISSR) and Simple Sequenced Repeat (SSR) markers have revealed a wide genetic diversity among *L. mutabilis* lines (Chirinos-Arias *et al.*, 2015; R. Galek *et al.*, 2017). In some cases, the variation illustrated by the analysis of genetic distance did not match the differences defined by morphological markers, suggesting that molecular markers other than ISSR and SSR may be more useful (R. Galek *et al.*, 2017).

Molecular and genetic tools available

At present, the availability of molecular resources for breeding of *L. mutabilis* remains scarce. The majority of molecular studies have so far focused on understanding *L. mutabilis* phylogeny. Initially, isozyme numbers revealed an affinity of *L. mutabilis* to the Old World species closer than that of any other North American species studied (B Wolko *et al.*, 1990). Later, the use of conserved chloroplast genes and internal transcribed spaces (ITS) highlighted the presence of an Andean group within the New World species (Ainouche *et al.*, 1999; Käss *et al.*, 1997; Michael Wink *et al.*, 1999).

TABLE 4| Suggested breeding traits for the improvement of *L. mutabilis*, goals and proposed strategies.

Breeding Traits	Goals	Proposed Strategies
Semi-determinate growth habit	• Determinate forms distinguished by medium-tall stems without lateral branches, resistant to lodgings and with early generative growth • Higher productivity and uniform maturation	• Identification of *rb* locus in *L. mutabilis* • Fixing the trait and breeding it into a stable variety
Environmental Adaptation	• Early maturing genotypes • Increased drought tolerance • Yield stability	• Selection of early maturing genotypes • Study the effect of vernalization on flowering time • Selection of genotypes based on photoperiod sensitivity • Investigation of drought tolerance strategies across genotypes • Breeding of homozygous lines
Alkaloid content	• Breeding of stable sweet varieties • Bitter/sweet lines	• Derive homozygous lines from "sweet" genotypes • Study the inheritance of alkaloid content in seeds • Study the translocation of QAs from source tissues to seeds • Target QAs transporters for the development of bitter/sweet lines
Seed color	• Seeds with uniform and heritable color (white)	• Select pure lines with uniform and heritable color patterns • Identify locus/loci responsible for color and patterns • Combine loci for high yield and white color
Proteins	• Identification and valorization of unique properties in *L. mutabilis* proteins • Increased production of ϒ- and β-conglutins; albumins and ferritin	• Identification of new valuable proteins • Elucidate biosynthetic pathways and functions of the different proteins
Oil	• Make *L. mutabilis* a dual-purpose crop for protein and oil	• Identification of accessions with low negative correlation between oil and protein • Elucidate relation between β-galactan and oil content in seeds

Only recently, the advent of nextRADseq technology has elucidated the area and timing of *L. mutabilis* domestication (Atchison *et al.*, 2016). Protein-based approaches have been carried out to determine seed storage protein composition in *L. mutabilis* and its differences between species

and lines (Santos *et al.*, 1997). Lately DNA based markers such as RFLP, AFLP, ISSR and RAPD have been used to assess genetic diversity between Lupinus species and have revealed a high intraspecific variation within *L. mutabilis* populations (Olczak *et al.*, 2001; Talhinhas *et al.*, 2003; Zoga *et al.*, 2008). A total of 113 SSR primers and 118 polymorphic InDel from *L. luteus* have been successfully used to characterize *L. mutabilis* genetically (Osorio *et al.*, 2018; Parra-González *et al.*, 2012).

Relative to other legumes, little genomic information is available for *L. mutabilis*. To date, even the number of EST's sequenced and submitted to the genomic databases remains very low (~65), and it mainly refers to molecular targets in ribosomal RNA (IGS and ITS) and other sequences used for taxonomic purposes (i.e. *rps16 gene*, submitted by (Keller *et al.*, 2017)). However, new developments in genomic technologies now provide a realistic opportunity to overcome the scarcity of genomic information and to hasten the identification of traits of interest. Over the last 15 years the limitations of approaches based on the identification of QTL derived from biparental crosses has shifted the focus towards association mapping in large panels of diverse genotypes. Genotype-by-sequencing (GBS) techniques can now provide thousands of single nucleotide polymorphism (SNP) markers at a much lower cost than earlier techniques, and they can be used to perform genotyping studies such as Genome Wide Association Studies (GWAS). In these studies natural populations hold the potential to replace recombinant populations in gene mapping and marker-trait associations (M.J. Iqbal *et al.*, 2012). With regard to *L. mutabilis*, GWAS could represent a possible approach to exploit the genetic resources of entire germplasm collections at once, while saving time and resources, exploiting multiple recombination events and considering the whole allele diversity. This kind of approach may serve as a foundation study and help to identify and establish valuable genetic markers for genomic selections, which will ultimately allow informed choices for further selection of breeding material and QTL analysis.

A wider selection of tools is available for *L. angustifolius* and *L. albus*, which have been more extensively studied in the past years. Genetic maps, BAC libraries, transcriptome and proteome assemblies, QTL and molecular markers for traits such as low alkaloids, flowering time and anthracnose disease resistance have been developed for these species and can potentially be exploited for *L. mutabilis* improvement (reviewed in Wolko et al., 2011; Abraham et al., 2019). Furthermore, the recent release of a high-quality genome draft for *L. angustifolius (951 Mb; 2n=40)*, and a high-quality chromosome-scale genome assembly for *L. albus* (451 Mb; 2n=50) represent a big support for the future whole-genome analysis of other lupin species, such as *L. mutabilis* (Hane *et al.* 2017). Similar 2C nuclear DNA contents were estimated in *L. mutabilis* (1.90 pg) and *L. angustifolius* (1.89 pg), suggesting that there might be a higher affinity between these two species (Naganowska *et al.*, 2003).

Applications and potential uses of *L. mutabilis*: much more than proteins

L. mutabilis emerges as a human health food and food additive, but its potential applications go far beyond food and target the utilization of the whole plant. *L. mutabilis* seeds represent an important and versatile source of proteins. Once debittered, the seeds can be directly consumed as a snack, or as an ingredient of many products and meals. In the Andean region they are traditionally used in soups, stew and salads or as raw material for preparing flour, milk and

margarine (Falconí, 2012). Like soybean, lupins also have important applications as food ingredients in many products: lupin flour, protein concentrate and protein isolate display physical and functional properties which are very valuable to the food and chemical sector (F.E. Carvajal-Larenas *et al.*, 2016). These derivatives can be used as base for meat alternative or replacers, as an egg replacement, as a bread improver, as an emulsifier and to increase the nutrient content of many products. After protein extraction, the large amount of dietary fibre still available (up to 40% of seed mass in *L. angustifolius*) can find application as prebiotic and human food ingredient in the production of fibre-enriched baked goods (Clark *et al.*, 2002; S.C. Smith *et al.*, 2006). The oil, characterized by a high nutritional value, also represents an attractive product for both nutraceutical and cosmetic purposes. Furthermore, pharmaceutics uses have also been described. *L. mutabilis* intake has been proven to reduce blood glucose and insulin levels, representing a valid alternative for treating hyperglycaemic diseases (Fornasini *et al.*, 2012). In the medical field, QAs also have an important role due to multiple properties such as anti-arrhythmic, anti-inflammatory, diuretic and hypotensive effects among others (Bunsupa *et al.*, 2012). In addition QAs can also find application in agriculture as a bio-stimulant increasing growth and yield of other crops (Przybylak *et al.*, 2005), as antibacterial agents (Romeo *et al.*, 2018) or as biocidal agents replacing synthetic toxins (Bermúdez-Torres *et al.*, 2009). Similarly a Blad-containing oligomer (BCO), a bioactive subunit of a polypeptide oligomer termed Blad (Banda de Lupinus albus doce) isolated in young cotyledons of Lupin spp. as a breakdown product of β-conglutin catabolism, has been recently introduced in the market as a novel fungicide against both human and phytopathogenic fungi, confirming the multiplicity of resources offered by this plant. BCO also act as plant bio stimulant and exhibits bactericide activity especially towards Gram+ bacteria (Carreira *et al.*, 2018). Beyond this, *L. mutabilis* can also be used as a fodder species. In the Andean area debittered seeds are used to feed pigs, sheep and poultry (Cremer, 1983). However, the optimal use of the plants for feed purposes would be as a green-fodder or silage, as debittered seeds are more profitable for food applications. Uses as silage or hay for livestock feed are mentioned in the literature, but its composition and nutritional value remain unstudied (Sherasia *et al.*, 2017). Similarly to other legumes, *L. mutabilis* is also able to assimilate atmospheric nitrogen and leave appreciable amounts in the soil as post-harvest residues of up to 400 kg ha^{-1} N (Adomas *et al.*, 2015; Brücher, 1989). Yet, it could prove more profitable to turn *L. mutabilis* biomass residues into bio-based products and energy sources, due to the boost in biomass demands in Europe.

Future prospects

Compared to many other pulses which dominate our agriculture (i.e. pea, lentil, fava bean), the domestication history of *L. mutabilis* appears very short and fragmented between Europe, South America and Australia. Even though global holdings of *L. mutabilis* represent a plethora of genetic resources, this source remains under-utilized and very often inaccessible. In addition, the lack of refined biotechnological methods in genetics, molecular cytogenetics or tissue culture, has limited the possibility of exploiting natural variability and performing distant crosses and haploidization of breeding material. The repeated use of a limited set of genetic resources in hybridization programs and the limited pre-breeding efforts account for a narrow genetic basis. Base broadening through mutation and hybridization -the main methods used so far- is a very slow process, taking many years before pure lines can be achieved. The coupled use of germplasm resources and modern approaches to broaden the genetic basis could now aid the introgression of desirable adaptive traits for specific environments, which are essential to develop *L. mutabilis*

into a valuable crop outside the Andes. The selection of genotypes adapted to specific latitudes and day lengths appear fundamental for farmers both in the Andes and in other parts of the world. Indeterminate growth habit and alkaloid content still represent a main limitation, but sweet lines and determinate forms with early maturation have been generated (R Galek *et al.*, 2007) (Table 3). A major effort is now required to fix these traits and breed them into stable variety for agricultural purposes. Currently, promising *L. mutabilis* lines are being screened throughout Europe aiming at the development of varieties adapted to European farming conditions within the next 10 years. Future work should focus on the development of bitter/sweet lines and on the promotion of different end-uses for proteins, oil and alkaloids which can contribute to increase the value of the crop in the near future. In this regard, studies combing genetic and multi-environment dataset will be important to unravel the genetic control of valuable traits. Further implementation of genomic selection and marker-assisted selection, will play a key role in speeding up breeding processes.

In spite of limitations, there remains enormous potential for the introduction of *L. mutabilis* as a protein crop. Its cultivation constitutes an important opportunity to provide a substantial source of protein through low input farming, both in the Andes and elsewhere in the world. In this regard, the potential of enhancing marginal lands production while contributing to the diversification of the protein market, righteously places *L. mutabilis* in the European agricultural system. Hence, *L. mutabilis* plays a major role on the protein transition scene, where plant based proteins will gradually replace animal proteins. Pivotal to achieving this aim are breeding programs focused on ensuring economic viability and consumer acceptance of the crop. Germplasm resources should be used together with conventional and molecular tools to unlock the genetic potential of *L. mutabilis* and secure it as a promising (new) protein crop. Finally, *L. mutabilis* represents a source of important traits for introduction into major lupin species or other legumes to aid their adaptation in a rapidly changing climate. Further research on this species can also provide valuable insights into important processes like protein and oil production in seeds or regulation of alkaloid content.

3

Diversity and Agronomic performance of *Lupinus mutabilis*: Germplasm in European and Andean Environments

Agata Gulisano [1], Sofia Alves [2], Diego Rodriguez[3], Angel Murillo[3], Bert-Ian van Dinter[4], Andres F. Torres[5], Milton Gordillo-Romero[5], Maria de Lourdes Torres[5], João Neves Martins[2], Maria-João Paulo[6] and Luisa M. Trindade [1]

[1] Wageningen University & Research Plant Breeding, Wageningen University, Wageningen, Netherlands,
[2] DRAT, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, Lisbon, Portugal
[3] INIAP, Estación Experimental Santa Catalina, Quito, Ecuador,
[4] Vandinter Semo, Scheemda, Netherlands,
[5] Plant Biotechnology Laboratory, Universidad San Francisco de Quito, Quito, Ecuador,
[6] Wageningen University and Research Biometris, Wageningen Research, Wageningen, Netherlands

Published as: Gulisano A, Alves S, Rodriguez D, Murillo A, van Dinter B-J, Torres AF, Gordillo-Romero M, Torres MdL, Neves-Martins J, Paulo M-J and Trindade LM (2022). "Diversity and Agronomic Performance of Lupinus mutabilis Germplasm in European and Andean Environments", **Frontiers in Plant Science**, doi: 10.3389/fpls.2022.903661.

Abstract

The introduction of *Lupinus mutabilis* (Andean lupin) in Europe will provide a new source of protein and oil for plant-based diets and biomass for bio-based products, while contributing to the improvement of marginal soils. This study evaluates for the first time the phenotypic variability of a large panel of *L. mutabilis* accessions both in their native environment and over two cropping conditions in Europe (winter crop in Mediterranean region/ summer crop in North-Central Europe), paving the way for the selection of accessions adapted to specific environments. The panel of 225 accessions included both germplasm pools from the Andean region and breeding lines from Europe. Notably, we report higher grain yield in Mediterranean winter-cropping conditions (18 g/plant) than in the native region (9 g/plant). Instead, North European summer-cropping conditions appear more suitable for biomass production (up to 2kg/plant). The phenotypic evaluation of 16 agronomical traits revealed significant variation in the panel. Principal component analyses pointed out flowering time, yield and architecture related traits as the main factors explaining variation between accessions. The Peruvian material stands out among the top yielding accessions in Europe, characterized by early lines with high grain yield (e.g. LIB065, LIB072, LIB155). Bolivian and Ecuadorian material appears more valuable for the selection of genotypes for Andean conditions and for biomass production in Europe. We also observe that flowering time in the different environments is influenced by temperature accumulation. Within the panel, it is possible to identify both early and late genotypes, characterized by different thermal thresholds (600-700 °C and 1000-1200 °C GDD, respectively). Indications on top-yielding and early/late accessions, heritability of morpho-physiological traits and their associations with grain yield are reported and remain largely environmental specific, underlining the importance of selecting useful genetic resources for specific environments. Altogether, these results suggest that the studied panel holds the genetic potential for the adaptation of *L. mutabilis* to Europe and set the basis for initiating a breeding program based on exploiting the variation described herein.

Keywords: Andean lupin, phenotypic diversity, germplasm characterization, breeding, grain yield, biomass, vegetative development

Introduction

Lupinus mutabilis, also known as tarwi, pearl or Andean lupin, is a high protein legume, native to the Andes and currently cultivated almost exclusively in Ecuador, Peru and Bolivia (Mercado *et al.* (2018). This species is characterized by the highest grain quality of all cultivated lupins. Its seeds have a protein and oil content similar to that of soybean, 44% dw and 18% dw, respectively, but are also characterized by the absence of starch and the presence of most essential amino acids, including methionine and cysteine (Gulisano *et al.*, 2019c). In the past 20 years, *L. mutabilis* has gained increasing interest in Europe, as it could represent a superior alternative to the current plant-based sources of protein and oil, such as soy or pea. The potential for the successful introduction of *L. mutabilis* to Europe is high as the crop is adapted to low input farming and temperate climatic conditions in the Andean region where it originates. Similarly, *L. mutabilis* could contribute to the development of sustainable and competitive biomass industries by increasing biomass supply from marginal lands.

L. mutabilis is the only economically important species of the genus *Lupinus* that originates from the New World (South America). *Lupinus albus*, *L. luteus* and *L. angustifolius* all originated from the Mediterranean region (Bogdan Wolko *et al.*, 2011) and are the most cultivated worldwide (Abraham *et al.*, 2019). However, old world lupin species are still characterized by relatively low and unstable yields, producing at their best between 2.5 and 4 t/ha and having a lower content of protein (34-42 %) and oil (5-11 %) in the seeds than *L. mutabilis* (Abraham *et al.*, 2019; Gulisano *et al.*, 2019). The domestication of lupins in the Old and New Worlds was completely independent but followed similar patterns, involving phenotypic changes towards non-shattering pods, permeable seed coats, and large seeds. Demographic analysis suggests that *L. mutabilis* was domesticated in northern Peru around 2600 years ago, as it split from the wild progenitor *L. piurensis*. *L. mutabilis* was derived from a small subset of the ancestral population which then went through a classical domestication selection process and a subsequent rapid population expansion as it became cultivated across the Andes (Atchison *et al.*, 2016). The presence of island-like habits and diverse ecological opportunities in the Andean region has led to exceptional rates of diversification in the whole Andean lupin clade (C. Hughes and Eastwood, 2006). Similarly, *L. mutabilis* is characterized by a high phenotypic diversity that allowed its adaptation to a wide range of altitudes and microhabitats as it expanded from Colombia to the north of Argentina.

At present, *L. mutabilis* remains an understudied crop, inadequately characterized and underutilized. The expansion of its cultivation has been strongly limited by the presence of toxic alkaloids in its seeds and low yields. The potentially high alkaloid levels do not represent an insurmountable barrier since low alkaloid genotypes are already available (reviewed in Gulisano *et al.*, 2019). The main hindrance to the establishment of *L. mutabilis* as a crop in Europe is primarily the species' low productivity and long vegetation periods. Previous studies have pointed out the importance of breeding for a better plant architecture and early maturity in order to address these bottlenecks (Caligari *et al.*, 2000). In fact, suboptimal yields are mainly caused by the indeterminate growth habit of the crop, which favors vegetative growth at the expense of seed production. Indeterminate growth is an undomesticated characteristic of many grain legumes and remains a major challenge for legume breeders. The ability to prolong indefinitely the vegetative phase after the onset of flowering is expressed in lupin species through the production of successive orders of branches throughout the cropping season. Indeterminate growth habit has been overcome in *L. albus, L. luteus* and *L. angustifolius* through the selection of

spontaneous or induced mutants with a determinate or semi-determinate growth habit (Gorynowicz *et al.*, 2014; Bogdan Wolko *et al.*, 2011).

The study of the genetic variation present within germplasm collections can support breeding programs in defining useful genetic resources, adaptation strategies and adaptive traits. The genetic diversity present in *L. mutabilis*, notably its Andean clade, can provide ample resources for breeding programs aimed at enhancing its commercial potential and successful introduction to new production systems. South American institutions started gathering *L. mutabilis* germplasm in 1947, and currently hold the largest and most diverse collections (including more than 3,000 different genotypes) in the gene banks of Peru, Ecuador, and Bolivia. However, a comprehensive study of *L. mutabilis* germplasm from the Andes and its performance across different environments is still lacking. Accordingly, preliminary evaluations of the performance of *L. mutabilis* in Europe have been carried out using a limited set of lines (resulting from mutations and repeated selection) across very similar environments (Caligari *et al.*, 2000; R Galek, 2010; R. Galek *et al.*, 2017; Guilengue, Alves*, et al.*, 2020). These studies have highlighted the availability of relevant breeding traits within the small panel of investigated genotypes (Guilengue, Alves*, et al.*, 2020), thus underscoring the importance and necessity for a large scale evaluation of *L. mutabilis* germplasm and its performance across different environments in Europe.

In this study, we address the aforesaid lack of evaluation of wide germplasm collections in different agroclimatic areas, both at a transnational and transcontinental level. We evaluated the largest collections of *L. mutabilis* accessions under study, comprising accessions from the Ecuadorian germplasm bank of the Instituto Nacional de Investigaciones Agropecuarias (INIAP) and European breeding programs. This large panel of 226 genetically diverse accessions was evaluated for its agronomic performance both in the native environment (Ecuador) and over the two potential cropping conditions in Europe. Field trials were set up to test the genotypic panel's suitability to serve as a winter and summer crop, in the Mediterranean and North/Central Europe respectively. Scoring of morphological, phenological and yield related traits provides valuable information for germplasm users and serves as basis for the classification of accessions adapted to specific environments and suitable for seed and biomass production.

Material and Methods

Plant Material

A panel of 225 genetically diverse *L. mutabilis* accessions was used in this study to investigate phenotypic variability within this species and to test *L. mutabilis* in different environments (see Figure 1). A large part of the panel comprised 201 accessions selected from the INIAP germplasm bank, and included land races, varieties and wild material collected across Ecuador (96), Peru (64), Bolivia (15) and 8 lines donated to the collection by the Byelorussian Agricultural Academy. The origin of 18 accessions remains unknown, due to the lack of passport data for this germplasm resources. The panel here described will be further referred to as GWAS panel. This panel was evaluated in four field trials, including one location in Ecuador and three locations in Europe, representing an example of cultivation in the native environment, winter-Mediterranean and summer-North/Central European cropping conditions. Two cultivars were included in the study as reference: I-450 ANDINO (INIAP, 1999) was included in the Ecuadorian field trial, while Inti (Chile, 1980) was included in the European field trials. Additionally, 24 *L. mutabilis* lines developed in breeding programs in Europe were evaluated in the European field trials. These lines were provided by the Instituto Superior de Agronomia (ISA, Lisbon, Portugal) and the

Julius Kühn-Institut (JKI, Quedlinburg, Germany) and are potentially better adapted to European conditions. A more detailed listing of all the accessions included in this study is available in Supplementary Table 1.

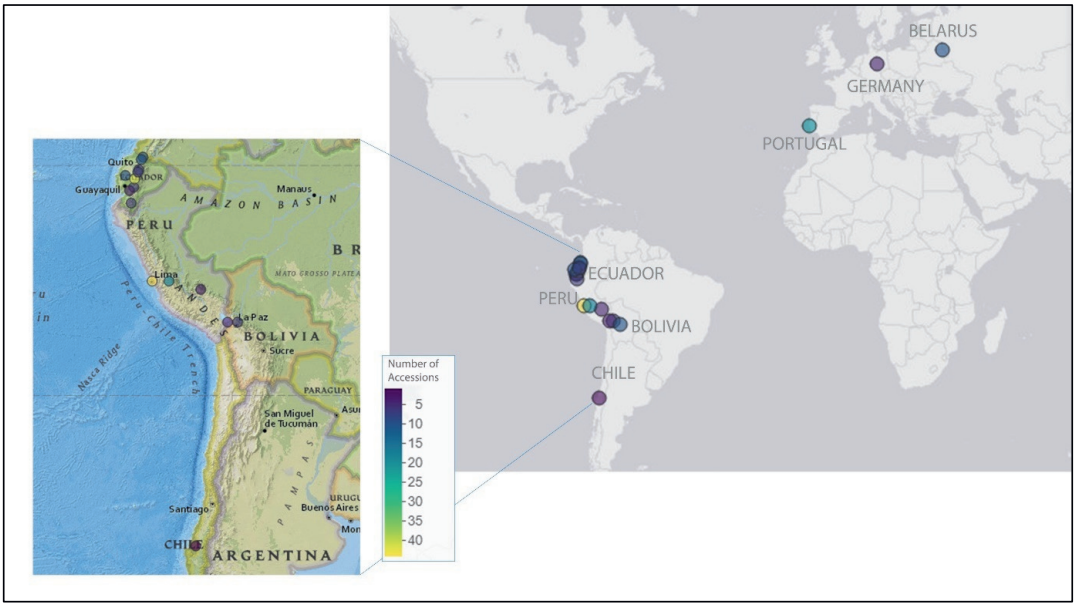


FIGURE 1| Geographic distribution of the 225 Lupinus mutabilis accessions included in the panel under study. Dots on the map indicate the provenience of the accessions. For the breeding material, dots indicate the institutes developing and providing the lines for our collection (ISA, Portugal; JKI, Germany; Byelorussian Agricultural Academy, Belarus). For the accessions from South America (donated by INIAP), dots indicate the provenience of landraces, varieties, and wild material collected along the Andean region, when these data were available. Maps were generated using mapview package in R (v. 2.11.0, Appelhans et al., 2022).

Field characterization

The effects of genotype, environment, and genotype by environment (GxE) interactions on the phenotypic variation of morphological traits were assessed on four locations, one in Lisbon in 2018 (PT), two in The Netherlands, respectively in Scheemda in 2019 (NL-Sc) and in Winschoten in 2020 (NL-Wi) and one in Cotopaxi, Ecuador in 2020 (EC). In Europe, field trials were grown during the Winter for the Portuguese site, and during the Summer in The Netherlands, in clay soil in 2019 and sandy soil in 2020. In Ecuador, *L. mutabilis* accessions were sown in December 2019 and harvested in June 2020, following local cultivation practices. In all locations, plants were cultivated under rain-fed conditions and without the aid of any fertilization (see Table 1). In Europe, a randomized complete block design in three replicates was used. The experimental units were plots including 20 plants (4 rows with 5 plants per row), at a distance of 30cm x 30cm. Phenotyping of morphological and phenological parameters was conducted only on the 6 central plants of the plots (biological replicates). In Ecuador, an Alpha Lattice design (40x5) in three replicates was used. The experimental units were rows of 2 meters, with 10 sowing spots at a distance of 20 cm, and 80 cm distance between plots. Phenotype was scored on the 5 central plants. Quantitative traits related to plant morphology, phenology and agronomic performance were measured in all the locations. Traits to phenotype were adapted from the IBPGR descriptors

(IBPGR, 1981) based on the feasibility of their scoring on large scale and keeping in mind the main breeding goals for *L. mutabilis*, namely earliness, plant architecture, grain and biomass yield.

TABLE 1| Environmental characteristics of the four field trial locations during the respective growing season

		EUROPEAN ENVIRONMENTS		
Trial name	**EC**	**PT**	**NL-Sc**	**NL-Wi**
Location	Cotopaxi, Ecuador	Lisbon, Portugal	Scheemda, Netherlands	Winschoten, Netherlands
Coordinates	0°55'35.1"S 78°40'07.4"W	38°42'33.5"N, 9°11'00.5"W	53°09'60.00" N, 6°57'59.99" E	53° 10' 11.346, 7° 2' 56.096"
Altitude (m)	2948	60	-1	3
Average Temperature (°C)	13.6	15	14	15
Total Precipitation (*mm*)	854.7	260.1	328.7	408.8
Day Length (*hours*)*	12	10→ 14.2	14 → 10.44	14 → 10.44
Maximum day Length (*Hours*)	12	14.2	16.5	16.5
Average relative humidity (%)	86%	74%	77%	75%
Average Wind speed (m/s)	5.2	21.7	4.24	15
Growing Season	December 2019-June 2020	November 2018-May 2019	April-October 2019	April-October 2020
Soil Type	Sandy	Clay	Clay	Sandy

Morphological measurements included plant height (scored at harvest), fresh biomass (fresh weight of the total biomass harvested, after eliminating the root system), number of branching orders (0= main stem only, 1= main stem and first branching order, etc.) and number of pods and seeds present respectively on the main stem (MS), on the first branching order (FO) and on the remaining branching orders (RO), as illustrated in Figure 2. In Portugal, due to the restricted architecture of the plants, scoring of pods and seeds was divided only over main stem production and the remaining part of the plant (mainly first order branches). Phenology of the accessions was investigated by scoring germination and flowering time (as number of days from sowing) and calculating growing degree days (GDD) accumulation using 4 °C as baseline temperature (as in Hardy *et al.*, 1998). Due to the impossibility of collecting data during Covid restrictions, measurements of phenological traits in Ecuador were not possible. Finally, agronomic performance assessment included measurements of total number of pods and seeds harvested, 100 Seeds weight (dry) and aboveground biomass fresh weight (g). Seed yield was measured per plant at harvest after drying the seeds at room temperature (dry weight, g/plant), while vegetative yield (dry weight, g/plant) was estimated as the difference between the total amount of biomass harvested (dry weight) and the seed yield per plant.

Statistical analysis

Single-site analysis were conducted in each environment, using the R package *statgenSTA* (van Rossum *et al.,* 2021). A linear mixed model with spatial correction was fitted for each trial separately and used to estimate adjusted means. The model fitted with R package *statgenSTA* (v1.0.8) included fixed effect for genotype, fixed effect per block and an extra spatial component obtained using SpATs, which uses 2-dimensional smoothing with P-splines as described in Rodríguez-Álvarez *et al.* (2018). By fitting a smooth surface through joint modelling of additive one-dimensional trends plus interactions between trends in the row and column directions, this method can account for all sources of continual environmental variation and explicitly model both large-scale and small-scale spatial dependences, leading to an improvement in precision and predictions of genotypic values (Rodriguez-Alvarez *et al.*, 2018; Velazco *et al.*, 2017). Adjusted means, Best Linear Unbiased Estimators (BLUEs), were obtained from the same model for each line in each environment. BLUEs were used to perform the Principal Component analysis per site using *FactoMine*R package and to calculate phenotypic correlation between traits using the *cor* function of base R. For the estimation of heritability, the same mixed model fitted with *statgenSTA* was used but treating genotypes as random factors.

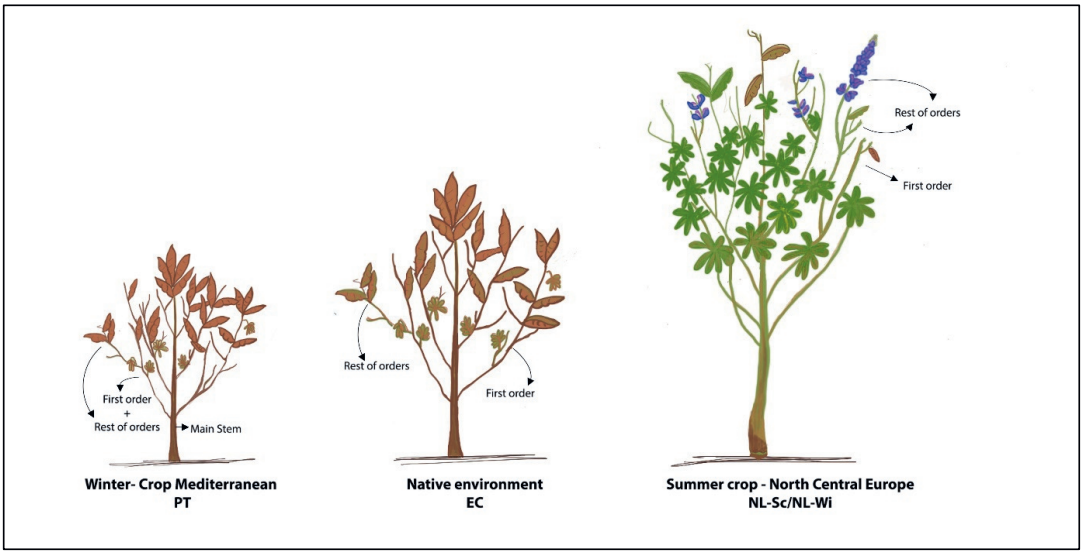


FIGURE 2| Morphological development of *Lupinus mutabilis* in the tested environments. At harvest, in PT and EC, plants reached full maturity, were dry, and their architecture was generally restricted to two branching orders (semi-determinate growth habit). In NL, at harvest, plants were highly branched, still switching between vegetative and reproductive phases and therefore holding floral buds and pods often no economic value (indeterminate growth habit). In evidence, the different architectural parts of the plant are the main stem (MS), the first branching order (FO), and the rest of the branching orders (RO). In PT, production on the first order and the rest of the orders were scored together.

Results

Cropping conditions highly impact morphology and yield of *Lupinus mutabilis*

A large and diverse panel of *L. mutabilis* accessions was used to test the agronomic performance of this species over three different cropping conditions, two in Europe and one in the Andean region (Ecuador), the native environment of this species. In Europe, it has been tested as a winter crop in the Mediterranean area (Portugal) and as a summer crop in North-Center Europe (The Netherlands) both on sandy and clay soils. The overall performance of *L. mutabilis* showed extreme variation between the different environments tested in Europe, and between European and Ecuadorian environments, in terms of yield, vegetative development and phenology (see Figure 1 and Table 2). In terms of grain yield, the best performance was observed in Portugal (on average 18.1 g/plant), followed by Ecuador (9.2 g/plant) and the Dutch site on sandy soil with 6.6 g/plant. The lowest yield was measured in the field trial in The Netherlands with clay soil (seed yield of 2.5 g/plant on average). This agronomic variation can be explained by the different climatic and soil conditions characterizing each environment. When growing on clay soil in winter-Mediterranean conditions (Portugal), *L. mutabilis* stayed generally smaller than in Ecuador, reaching a height of 60 cm and producing the lowest amount of biomass (116 g/plant). Low precipitation and high temperatures after flowering, contributed to seed maturation in this environment and restricted biomass accumulation due to drought. Also worth noting, in Portugal, plants were slightly more branched than in Ecuador and developed more often a second branching order (Orders= 1.6). On the other hand, summer cropping conditions in The Netherlands prompted a higher vegetative development and indeterminacy, that severely impacted grain yield. With higher precipitation, *L. mutabilis* accessions grew taller (~ 90 cm) and developed more than two branching orders (1.88-2.65), generating an increase in production of above ground biomass. In the Dutch site with clay soil (NL-Sc), the performance was generally poorer, and plants were characterized by both the lowest grain and biomass yield. Short days in Portugal did not significantly affect flowering time (110 days), which was in line with general flowering time observed in Ecuador under a fixed photoperiod. Conversely, long days in The Netherlands remarkably shortened flowering time by about a month (to 80-90 days). However, in both European environments, it was possible to distinguish earlier and late flowering accessions characterized by a different growing degree days (GDD) accumulation. In Portugal early accessions had a GDD of about 600 °C and flowered before 105 days after sowing, while late accessions flowered between 116-127 days with a GDD of about 1000 °C. In The Netherlands, some accessions were very early and flowered between 73-80 days after sowing (GDD ≈700 °C), while a higher proportion was late and flowered between 90-104 days (GDD ≈1200 °C).

High variability in the studied panel for several morphological traits

Phenotypic evaluation of the *L. mutabilis* panel revealed significant variation in many of the measured traits, confirming the presence of significant variability within the studied panel of accessions. Raw data were first corrected for spatial variation within each trial, obtaining BLUEs values for each trial that were used in further analysis. Spatial variation is common in field trials, and accounting for it increases the accuracy of estimated genetic effects (Selle *et al.*, 2019). After the correction, the variability observed was decomposed into genetic, spatial and residual variance and heritability values were estimated. Table 2 summarizes BLUEs and heritability values estimated per each trait in each of the tested environments. A large range of variation for the scored traits pointed out a high variability between accessions, especially for yield related traits such as production of pods and seeds on the first branching order and on the rest of the plant

(see Table 2). High heritability for plant height (0.5-0.9) and seeds and pods production on the main stem (0.4-0.8) was observed in all the field trials. Flowering time, which is a genetically determined trait, was highly heritable (0.9) both in PT and NL-Wi. The low value of heritability reported for flowering in NL-Sc is explained by the high amount of variance associated to residual and spatial variation (see Suppl. Table 2), probably caused by the large influence that climatic conditions had on flowering in this location. Conversely, the estimation of higher heritability for total number of seed and vegetative yield in the environments where the mean estimates for these traits were lower (respectively Portugal and the Netherlands, see Table 2) can be explained by the smaller range of phenotypic variation assessed and associated to spatial variation (see Suppl. Table 2). Limiting environmental conditions for these traits, such as drought for vegetative yield in PT and abundant precipitations for seed yield in NL, contributed to even out phenotypic variation within accessions and resulted in a higher variation between accessions, resulting in higher heritability estimates (see Supp. Table 2).

TABLE 2| Mean, range of variation, and broad-sense heritability (H^2) for the blues values of morphological and phenological traits were assessed across the different field trials.

	Native environment			Winter crop-Mediterranean			Summer crop - north-central Europe					
	EC			PT			NL-Sc			NL-Wi		
	Mean	Range	H^2	Mean	Range	H^2	Mean	Range	H^2	Mean	Range	H^2
Germination time *(days)*	-	-	-	10.8	[8.06, 16.7]	**0.51**	-	-	-	17.8	[11.6, 32.0]	**0.78**
Height MS *(cm)*	73.6	[24.2, 95.9]	**0.72**	59.9	[27.5, 104]	**0.86**	90.6	[41.3, 208]	**0.56**	92.7	[47.5, 121]	**0.57**
Flowering time *(days)*	-	-	-	110	[91.0, 127]	**0.93**	90.8	[75.6, 113]	0.24	82	[65.8, 98.6]	**0.87**
Fresh biomass *(g)*	-	-	-	116	[4.45, 240]	**0.68**	507	[132, 1910]	0.19	750	[93.7, 2010]	0.32
Branching Orders	1.09	[0.225, 1.97]	0.33	1.6	[0.798, 2.18]	**0.44**	1.88	[0.868, 2.78]	**0.46**	2.65	[1.34, 3.93]	0.07
Pods MS	8.74	[4.76, 17.5]	**0.48**	12.8	[5.20, 26.2]	**0.42**	2.84	[0.023, 13.4]	**0.83**	2.84	[0.023, 13.4]	**0.46**
Pods FO	7.17	[0, 27.2]	0.18	24.1	[8.27, 48.9]	**0.43**			**0.54**	11.9	[0, 44.3]	0.39
Pods RO	0.29	[0, 7.07]	0.07				3.38	[0, 18.3]	0.39	5.51	[0, 47.3]	0.12
Pods T	16.1	[2.34, 41.3]	0.23	36.1	[19.7, 70.7]	**0.41**	7.79	[0, 20.7]	**0.69**	19.7	[5.29, 62.1]	**0.58**
Seeds MS	25.2	[10.9, 56.0]	**0.46**	42.3	[17.0, 67.4]	**0.57**	13.9	[0.287, 39.7]	**0.83**	19.9	[5.02, 51.8]	**0.53**
Seeds FO	16.2	[0, 73.7]	0.15	54.7	[22.3, 108]	0.33	14.7	[0, 55.3]	**0.55**	18.3	[0,120]	**0.62**
Seeds RO	0.374	[0, 14.9]	0.07				6.32	[0, 26.5]	0.38	0.856	[0, 20.8]	-
Seeds T	41.6	[3.74, 119]	0.21	95.4	[50.9, 158]	0.36	26.4	[0.839, 71.7]	**0.70**	41.7	[5.30, 175]	**0.64**
100Seed weight *(g)*	23.2	[9.70, 30.1]	**0.67**	25.3	[7.63, 59.9]	0.17	18.6	[8.13, 28.8]	**0.63**	42.1	[11.2, 145]	0.24
Seed yield *(g/plant)*	9.16	[0.3, 25.9]	0.15	18.1	[6.5, 32.5]	0.28	2.5	[0.5, 11.3]	0.32	6.57	[0.5, 16.5]	0.37
Vegetative yield *(g/plant)*	-	-	-	45.8	[15, 97.8]	0.54	133.6	[30.7, 643.11]	0.17	174.9	[13.5, 451.2]	0.21

Values in bold signify a relatively high heritability.

Relatedness between accessions and contributions of single traits to the total variance were estimated via a Principal Component Analysis for each trial. Flowering time, yield related traits (seed yield, vegetative yield) and architecture related traits (nr. of branching orders and nr. of seeds from the different branching orders) were the main factors contributing to Principal Components and explaining the larger fraction of variation between *L. mutabilis* accessions performance in each environment.

Variability in Ecuador is largely due to yield and plant architecture-related traits

In Ecuador (Figure 3A), the variability among accessions was explained mainly by the total number of pods/seed and the share of grain production on the first branching order in PC_1 (58 %) and by the share of production on the remaining branching orders in PC_2 (14.7%). In their native environment, the majority of *L. mutabilis* accessions was characterized by low scores for total seed yield primarily from the first branching orders (clustered on the left side of the graph, Fig. 3A). In opposition, clustered on the right side of the graph we find the accessions with higher grain yield, and few accessions characterized by a particularly high production on the secondary branching orders (bottom right) and on the main stem (top right). The five accessions with higher grain yield produced between 22-26 g/plant in this environment (listed in Table 3) and include both lines characterized by a higher yield on the main stem (LIB021) and first branching orders (LIB007) and lines characterized by a high grain production coming from higher branching orders.

Two different morphotypes are identified under European cropping conditions

In European environments (Fig. 3B, 3C and 3D), it is possible to observe a more defined clustering around germplasm pools of accessions with the same origin, that exhibit similar phenotypes. In Mediterranean winter cropping conditions (Fig. 3B), almost 80 % of the variation observed across the panel was explained by the first three principal components. PC_1 explained 39.2 % of the variation by opposing accessions with larger coefficients for traits such as fresh weight and seed yield from secondary branching orders on the right, to smaller accessions with a seed yield mainly concentrated on the main stem on the left. PC_2 explained 19.2 % of the variance, dividing on the vertical plane landraces collected across South America characterized by later and taller plants, from accessions resulting from previous breeding programs in Portugal, Belarus and Germany characterized by a higher production of seeds on the main stem. On the other hand, in both trials in The Netherlands (Fig. 3C and 3D) five dimensions were retained to explain 80% of the variance. Around 40% of the variation was already explained in the first dimension that opposed accessions based on their seed production concentrated on the main stem and first branching order. Instead, the second dimension divided accessions mainly on the basis of their vegetative development, opposing accessions characterized by high biomass yield and indeterminate growth (top) to smaller and more determinate accessions (bottom). Similarly to Portugal, we observed a higher vegetative development and lower seed yield for accessions from INIAP collection and higher seed yield accompanied by a restricted development for breeding lines developed for European environments. However, in this environment, accessions from Peru showed a certain proximity to breeding lines, leading in certain cases to even superior yields. This distinction was particularly evident in the trial on sandy soil (NL-Wi, Fig. 3D), where the presence of favorable conditions increased the discrepancy in performance between lines from European breeding programs and other germplasms material (more sparse points). Notably, the five top grain yielding accessions (listed in Table 3) differ in each environment. Similarly to

Ecuador, in Portugal top yielding genotypes include both lines characterized by a higher yield on the main stem and lines characterized by a high grain production coming from higher branching orders. It is important though to consider that in these environments the seed production scored as coming from the other orders (thus excluding the main stem production), is in fact coming only from first and second branching orders due to the limited branching observed in these conditions. Conversely in the Netherlands, where seed yield was distributed among more branching orders, the highest yielding accessions had a higher seed production on the main stem and on the first branching order. For the European trials, in Table 4 are listed the two top yielding accessions in terms of fresh biomass and vegetative yield for each environment and biomass data are reported also for the top grain yielding accessions presented in Table 3.

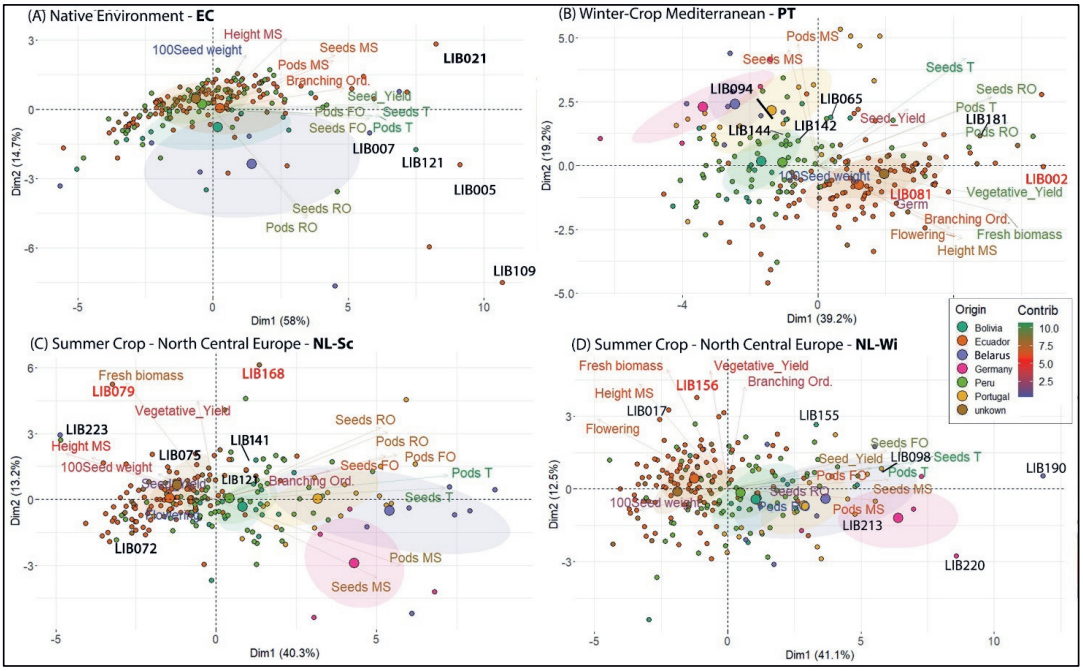


FIGURE 3| Principal component analysis of *L. mutabilis* collection, including 201 lines from the INIAP gene bank (Bolivia, Ecuador, Peru, unknown, and Belarus), Andino, and 24 lines from breeding programs in Europe (Germany and Portugal). Each biplot shows the PCA scores of the explanatory variables (as vectors) and individuals (as points) separately for each of the environments tested: (A) Ecuador, (B) Portugal, (C) NL-Sc and (D) NL-Wi. Individuals on the same side as a given variable should be interpreted as having a high contribution to it. The color of the explanatory variables (vectors) shows the strength of their contribution to each PC. The five most high-yielding genotypes for each location are indicated on the graph with a black label. The accessions with the higher biomass yield in European trials are indicated in red (reported in Table 4).

TABLE 3| BLUEs data for Seed yield (g/plant) and flowering time (days from sowing) of the five most yielding accessions in each environment.

	Ecuador				Portugal				Netherlands- Sc				Netherlands- Wi			
	Geno	**Origin**	**Seed yield**	**Flowering time**	**Geno**	**Origin**	**Seed yield**	**Flowering time**	**Geno**	**Origin**	**Seed yield**	**Flowering time**	**Geno**	**Origin**	**Seed yield**	**Flowering time**
1	LIB109	Ecuador	25.9	-	LIB065	Peru	32.4	102	LIB072	Peru	11.26	88	LIB155	Peru	16.5	68
2	LIB021	Ecuador	23.9	-	LIB142	Peru	26.7	107	LIB223	Peru	7.5	95	LIB098	Peru	15.3	79
3	LIB005	Ecuador	22.8	-	LIB144	Peru	26.6	105	LIB075	Ecuador	6.2	108	LIB220	Germany	14.2	71
4	LIB007	Ecuador	22.4	-	LIB181	unknown	26.3	115	LIB121	Bolivia	5.8	99	LIB213	Portugal	14.2	68
5	LIB121	Bolivia	22.1	-	LIB094	Peru	26.1	97	LIB147	Peru	5.7	84	LIB190	Belarus	13.5	69

TABLE 4| BLUEs data for fresh biomass yield (fresh weight of total aboveground biomass, g/plant) and vegetative yield (dry weight of stems and leaves, g/plant) in Europe. In order, the two higher yielding accessions in terms of biomass in each environment, followed by the five accessions with higher grain yield.

	Portugal				Netherlands- Sc				Netherlands- Wi			
	Geno	**Origin**	**Fresh Biomass**	**Vegetative yield**	**Geno**	**Origin**	**Fresh Biomass**	**Vegetative yield**	**Geno**	**Origin**	**Fresh Biomass**	**Vegetative yield**
1	LIB081	Ecuador	239.9	74.2	LIB079	Ecuador	1907.3	489.6	LIB017	Ecuador	2012.6	367.3
2	LIB002	Ecuador	227.6	89.4	LIB168	unknown	1430.8	424.2	LIB156	Ecuador	1759.1	423.4
(1)	LIB065	Peru	114.5	50.3	LIB072	Peru	417.2	100.8	LIB155	Peru	756.7	170.2
(2)	LIB142	Peru	111.7	53.7	LIB223	Peru	1136.9	434.2	LIB098	Peru	543.3	163.5
(3)	LIB144	Peru	110.8	33.7	LIB075	Ecuador	531.9	44.2	LIB220	Germany	684.1	126.6
(4)	LIB181	unknown	188.1	62.6	LIB121	Bolivia	426.6	135.3	LIB213	Portugal	521.2	100.2
(5)	LIB094	Peru	86.6	41.2	LIB147	Peru	293.2	71.1	LIB190	Belarus	328.3	121.5

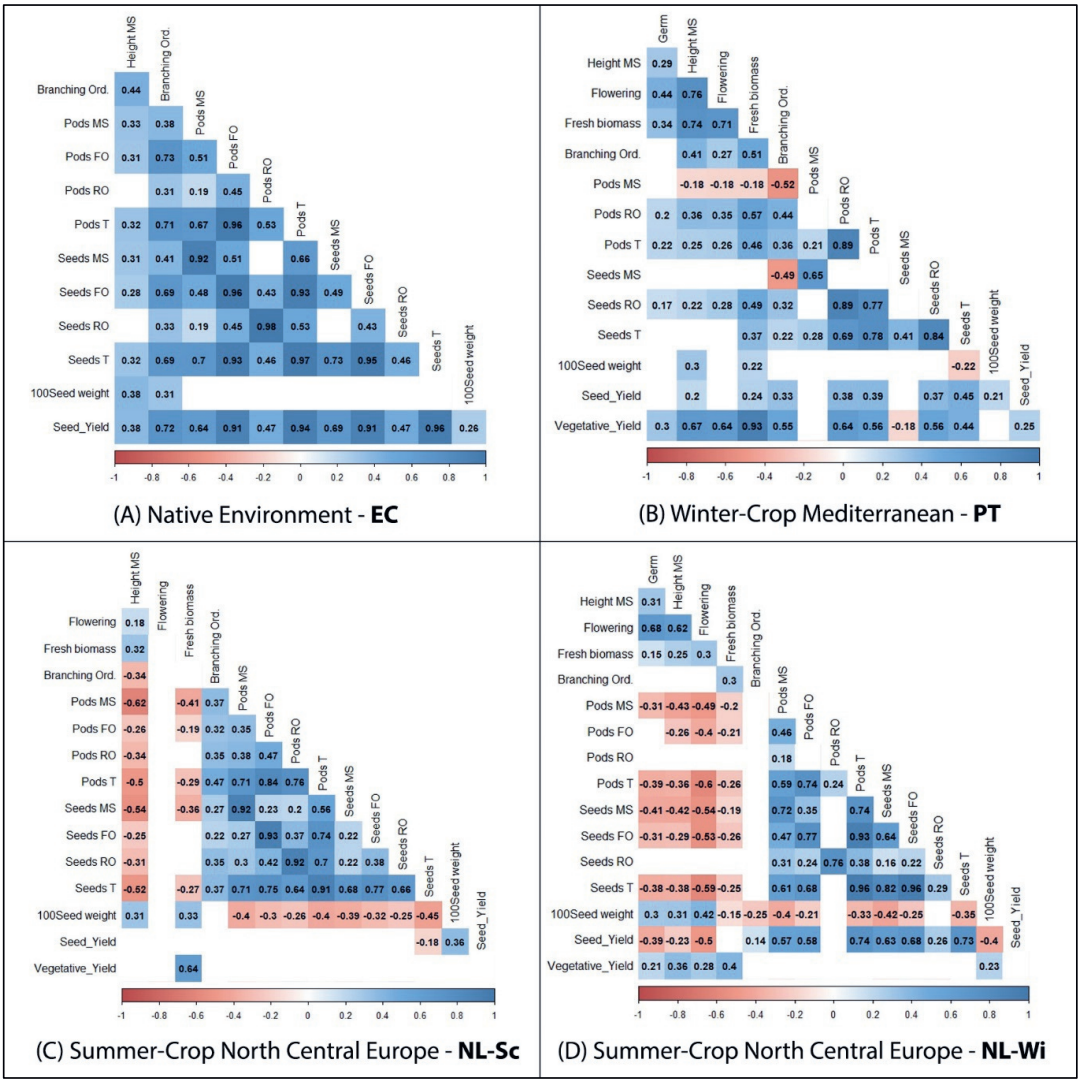


FIGURE 4| Correlation analysis between scored phenotypic traits (BLUEs data), presented per single field trial: (A) Ecuador, (B) Portugal, (C) NL-Sc, and (D) NL-Wi. Red and blue squares indicate significant positive or negative correlation (p-value < 0.01), whereas blank cells indicate no significant correlation (p-value > 0.01).

Prolonged vegetative development negatively affects grain production

BLUEs data were used to investigate correlations between traits (see Fig. 4). Significant correlations were found across trials indicating that the total number of seed produced were mainly correlated with the production on the first branching order (r = 0.77-0.96), followed by the production on the main stem (r = 0.41-0.82) and only to a lower extent with the production on the rest of branching orders (r = 0.3-0.66). Only in PT, where the count of seeds on the rest

of orders included seeds produced on the first branching order, the correlation with the total number of seeds was higher (r =0.84, Fig. 4B). Conversely, grain production is negatively correlated to vegetative development (height, fresh biomass) in North-Central European conditions while in PT only grain production on the main stem appears to be negatively correlated to an increasing number of branching orders. Furthermore, flowering time was highly correlated with vegetative development in PT where later-maturing accessions produced higher amount of biomass yield (r =0.64-0.71, Fig. 4B) and negatively correlated with seed yield in NL-Wi where later-maturing accessions had a lower seed yield (r =-0.5, Fig. 4D). In Ecuador, all correlations were positive.

Discussion & Conclusions

The importance of exploring genetic diversity for the development and selection of superior genotypes is the basis of many breeding programs. To our knowledge, this is the first study exploring the genetic variation and transcontinental performance of an extensive panel of 225 *L. mutabilis* accessions on four geographically distinct trials, including the native Andean environment and two potentially favorable cropping conditions in Europe. The panel of genotypes selected includes for the first time both a wide share of the germplasm collected in the Andean region (201 accessions) and a collection of lines derived from breeding programs taking place in Europe (24 lines). Until now, the lack of such in-depth wide-ranging studies has represented a limiting factor for the optimization of *L. mutabilis*, hindering its adoption on a commercial scale. Previous work has focused on assessing variation within small pools of genotypes mainly generated by induced mutations or as a product of crossing (Caligari *et al.*, 2000; Clements *et al.*, 2008; R. Galek *et al.*, 2017; Guilengue *et al.*, 2020; Galek *et al.*, 2016). As a result, the investigation of the performance and genetic variation available in wider germplasm collections has been highly recommended as a key approach in the generation of tools for crop improvement (González-Andrés *et al.*, 2007; Atnaf *et al.*, 2015; Mousavi-Derazmahalleh, Bayer*, et al.*, 2018; Annicchiarico *et al.*, 2010). Moreover, the study of germplasm collections across different environments can play an important role in identifying useful genetic resources for targeted breeding and in achieving adaptation to local climatologic conditions (Annicchiarico *et al.*, 2010).

Taking as a reference the behavior of *L. mutabilis* Andean germplasm in its native environment, we observe a very high variation in the agronomic performance of this species in contrasting environments in Europe. Furthermore, germplasm pools summarize a reasonably high portion of the observed variation, and are therefore of practical interest for identifying genetic resources which are likely to possess high adaptation to local conditions or other desired breeding traits. Interestingly, we observe similar range of diversity for Andean germplasm in the native environment while in European conditions a clear separation between the performance of Ecuadorian and Peruvian material is noted.

These findings reinforce the importance of evaluating large germplasm collections in contrasting environments representing major growing conditions and underline the importance of selecting useful genetic resources for a breeding program on the basis of landrace evaluation on the targeted environment, as also discussed by other authors (Annicchiarico *et al.*, 2010).

In Mediterranean- winter cropping conditions, higher temperatures and water scarcity highly affected the architecture of all the accessions, limiting crop development to two branching orders and increasing yield overall in the panel (>18 g/plant). These observations are in line with

previous studies reporting a high effect of temperature and water availability on the degree of determinacy of *L. mutabilis* in Mediterranean conditions, as plants tend to accelerate maturation under dry conditions (A Hardy *et al.*, 1997). Even though the environment has a main effect on the architecture of the crop, genotype per environment interactions remain relevant and point out the better adaptation of certain accessions to Mediterranean conditions. Landraces from Peru appear as promising germplasm material to increase grain yield in Mediterranean cropping conditions, as they outperform ISA lines selected specifically for this environment. We report four Peruvian landraces (LIB065, LIB142, LIB144 and LIB094) producing more than 26 g/plant of dry seeds, which could lead to grain yield close or above 3 t/ha, very similar to the values of *L. albus* grain yield growing under the same cropping conditions (Lo´ pez-Bellido *et al.,* 2000). These selected accessions are all characterized by very early flowering (97-106 days), short stature (60 cm) and generally by the development of a single branching order. The highest yielding accession, LIB065, reached a yield of 32.4 g/plant by having both a high seed production and a higher seed weight (30 g/100 seeds). The high performance of this accession can be explained by its origin in the Mantaro valley (Peru), an area characterized by a dry climate where crops are adapted to cultivation under rain fed conditions and drought as main limiting factor (Silva *et al.*, 2010). The precise origin of the other Peruvian accessions is not known, but they have likely been collected from the highlands of Peru, where ecotypes have been described as small, scarcely branched and early flowering (Doris Chalampuente-Flores *et al.*, 2021). Instead, ISA breeding lines exhibited a slightly lower branching degree than the Peruvian lines but their yields were lower, ranging between 7 g/plant (LIB202) and 25.3 g/plant (LIB204). Following LIB204, LIB214 (23.3 g/plant) and LIB212 (21.5 g/plant) appear to be the best ISA accessions for this environment, confirming their good agronomic performance in previous studies (Guilengue, Alves*, et al.*, 2020; Lazaridi *et al.*, 2020). Therefore, our results expands prior work pointing out the possibility of improving the performance of already selected lines for Europe by making use of the high breeding potential of germplasm from Peru. Notably, a later genotype of unknown origin and characterized by a higher degree of branching orders (LIB181) was also clustered within the genotypes with the highest yield. This supports prior hypothesis (Martins *et al.*, 2016) that in unstable climates, such as the Mediterranean one, plants with an indeterminate growth might have the advantage to allow a better compensation of yield when adverse climatic conditions affects pods set, therefore guaranteeing a higher yield.

In North-European summer cropping conditions, a more constant and higher precipitation prompted *L. mutabilis* vegetative growth throughout the entire cropping season, resulting in highly branched plants holding floral buds and pods of no economic value at harvest. As observed in other legumes, an indeterminate growth habit highly affects the productivity making the difference between potential productivity and actual yield striking (A. Sinjushin, 2015a). Hence, it is not surprising that accessions characterized by a semi-determinate growth habit (specifically, European breeding lines and Peruvian landraces), were the best performers in the Northern European environments. In fact, the determinacy characteristic of these accessions enforces a switch towards the productive phase, which guarantees a more stable seed yield in this environment. Notably, previous evaluation of ISA line LIB213 have reported very late flowering and low productivity in Mediterranean environments (Guilengue, Alves*, et al.*, 2020). Instead, in North European environments, we report relatively high yield and early flowering for this accession. A similar pattern was also observed for line LIB220 and LIB223. LIB220 - which has been selected as a semi-determinate growth type and has been characterized as non-suitable for cultivation in the Mediterranean area (Lazaridi *et al.*, 2020) showed a better adaptation in the north of Europe, in NL-Wi. LIB223, described in literature as a long life-cycle indeterminate accession in Mediterranean conditions (Lazaridi *et al.,* 2020), was actually within the most yielding

genotypes when encountering unfavorable conditions in North-European conditions (NL-Sc). Overall, the difference in grain yield across the two trials in The Netherlands was striking, and very large differences in seed yield were observed for many lines, including top yielding accessions. On average, yield on sandy soil exceeded the yield on clay soil by 65%. Lupins characteristically grow on well-drained acidic to neutral soils and are generally intolerant to waterlogging. Waterlogging is common in clay soils and has been found to affect root growth and plant development (Dracup *et al.*, 1998). However, the striking difference in grain yield can also be explained by the amount and distribution of rainfall. Studies on the effect of rainfall on the development and yield of blue lupin in temperate climates have shown that seed yield is dependent to the largest extent on the amount of rainfall in June and July, which are the periods of blooming and pod setting (Podlesny *et al.*, 2011). The scarcity of rainfall registered for this period in 2019 (2.25 times less than in 2020) might have therefore been the main reason behind the significant decrease in yield observed in NL-Sc (see Suppl. Fig. 1).

Considering these results, the use of Peruvian germplasm as basis for breeding programs aimed at increasing seed yield in *L. mutabilis* for European cropping conditions is recommended. However, the use of Ecuadorian and Bolivian germplasm material remains highly valuable for the generation of variation in breeding germplasm and for the selection of new genotypes adapted to Andean conditions. In Table 3, we present a selection of landraces collected between 2600 and 3000 m of altitude, which are best adapted to the conditions of Cotopaxi (at 2948 m) and produce more than 20g seed/plant. These grain yields are superior to the ones of selected varieties such as I-450 ANDINO (13 g/plant) or LIB091 (17 g/plant; reported as ECU-2658 in (Guaytarilla *et al.*, 2014)), and offer the possibility to further increase grain yield via breeding. Notwithstanding, our indications on useful genetic resources cannot be considered as conclusive because the lack of repetition in time of the trials prevents to assess the bias derived by genotype x year interactions. However, various features of the germplasm pools such as higher determinacy of Peruvian and European breeding lines, higher biomass production of Ecuadorian material and earliness of Peruvian and Bolivian landraces support previous findings (Schoeneberger *et al.*, 1982; Gross and Baer, 1981). Hence, these factors have implications for establishing breeding programs for specific locations and increasing the efficiency of genetic resource evaluation aimed at the identification of elite parental material.

Our results clearly indicate that grain yield and architectural traits are strictly correlated, as the highest yielding accessions had a semi-determined growth type, with the main part of the production coming from the main stem (20-50%) and the first branching order (50-80%). Our findings corroborate previous studies that observe high correlations between seed yield and production of pods and seeds in the main stem and first branching orders, concluding that the selection of semi-determinate accessions with restricted branching is pivotal to achieving high yield stability in this species (Caligari *et al.*, 2000; Clements *et al.*, 2008; Guilengue, Alves*, et al.*, 2020). Recessive genes responsible for the restriction of branching have been identified in several lupin species, and the inheritance of this character has been confirmed as monogenic recessive also for *L. mutabilis* (Römer, 1994). Notably, stem height and production of seeds and pods on the main stem appeared as highly heritable across trials, making them good targets for breeding. Although the use of two different experimental designs in Ecuador and Europe, estimations of highly heritable traits were fairly consistent across trials. An accurate estimation of genotype and environment effects, generally considered more precise on large trials when using incomplete designs (Kumar *et al.*, 2020), was here guaranteed also for the plots with a RCB design through the use of spatial correction. In agreement with other studies, we observed high heritability values for flowering time (Harzic *et al.*, 1996 1996). Flowering time was very important for adaptation

to both European cropping conditions and appeared negatively correlated to seed yield (r = -0.5). Similarly to what was observed in *L. angustifolius* (Reader *et al.*, 1995) and *L. albus* (Christiansen *et al.*, 2002), time to flowering was reduced when plants were exposed to increasing of the day length (The Netherlands, 70-90 days). Instead, in Portugal under short days, flowering time remained similar to that observed in the Andean region (90-120 days). Flowering time in other lupin species has been shown to be highly responsive to both daylength and temperature (Reader *et al.*, 1995; Kedar N. Adhikari *et al.*, 2012). Previous reports on photoperiodic sensitivity of *L. mutabilis* are contrasting (Hackbarth, 1961; Jacobsen and Mujica, 2008), but flowering time has been reported as highly responsive to the environment, with a predominant effect of temperature (Hardy *et al.*, 1998). Our results suggest that thermal accumulation might play an important role in flower initiation and classify early accessions with a GDD sum of 600-700 °C and late genotypes with a GDD of about 1000-1200 °C. Similar flowering range have been reported on a small subset of this panel propagated in Portugal in 1994-95 (A. Hardy *et al.*, 1998). Even though further studies are needed to clarify the mechanisms regulating flowering time in *L. mutabilis*, our findings indicate that a genotype specific amount of heat units above a base temperature are required to start flowering. Hence, these findings can be useful in the selection of accessions for specific environments. Early accessions are preferable in winter Mediterranean conditions owing to the shorter growing season and to the possibility of escaping terminal drought and heat stress, reducing loss in yield. In North-Europe, earlier accessions performed better under optimal conditions, however intermediate flowering resulted in higher yield when the conditions were suboptimal.

Altogether, our observations show that this *L. mutabilis* panel holds the genetic potential for the adaptation of this crop to European environments. The comparison between *L. mutabilis* agronomic performance in its center of origin and two different environments in Europe highlights potential grain yields up to 3t/ha in Mediterranean winter conditions, superior to the one achieved in the native region. The capacity of producing such a grain yield in poor soils and under rain-fed conditions, confirm *L. mutabilis* as a potent candidate for sustainable protein production in Europe, and a crop with potential for marginal land. Conversely, the agronomic performance observed in North European conditions confirms the findings of Caligari (Caligari *et al.*, 2000) who singled out indeterminate growth habit as the main factor limiting the yield of *L. mutabilis* in North- European field trials. This suggests that semi-determinate type remain preferred to ensure stable yields, whilst grain yield, earliness and plant architecture emerge as key traits for breeding of selected genotypes adapted to specific agroecological conditions. With our analysis we aim to provide the framework for initiating a *L. mutabilis* breeding program based on exploiting the available variation described herein. Finally, we report potential biomass yields up to 2 kg/plant of fresh biomass and to 0.5 kg/plant of dry vegetative material in North European cropping conditions suggesting that *L. mutabilis* could find application also as a source of lignocellulosic biomass, generating over 50 t/ha of agricultural residues. The valuable use of the whole plant could significantly support and enhance European bio-economy by supplying biomass from marginal lands requiring no additional inputs. Future work should focus on assessing the quality of the biomass and testing its possible application in different bio-based products.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.903661/full#supplementary-material

4

Investigating the potential of Andean lupin as a lignocellulosic feedstock for Europe: first genome-wide association study on *L. mutabilis* biomass quality.

Agata Gulisano[1], Annemarie Dechesne[1], Maria-João Paulo[2], Luisa Trindade[1]

[1] Wageningen University & Research Plant Breeding, Wageningen University, Wageningen, Netherlands,
[2] Wageningen University and Research Biometris, Wageningen Research, Wageningen, Netherlands
[3] INIAP, Estación Experimental Santa Catalina, Quito, Ecuador

Published (with modifications) as: Gulisano, A., Dechesne, A., Paulo, M.-J. and Trindade, L.M. (2022). "Investigating the potential of Andean lupin as a lignocellulosic feedstock for Europe: first genome-wide association study on L. mutabilis biomass quality", **GCB Bioenergy**, https://doi.org/10.1111/gcbb.13006

Abstract

The development of multipurpose crops will drive the transformation of agricultural "waste" into added-value products, helping to meet biomass demands without competing with food production or increasing environmental pressure. *Lupinus mutabilis*, has been proposed as a valid source of protein and oil for Europe, but also as a possible source of lignocellulosic feedstock for the biorefinery industry. In this study, the quality of *L. mutabilis* lignocellulosic biomass and its genetic architecture are investigated for the first time, using a panel of 223 accessions planted across three locations in two different European cropping conditions. Biomass quality was evaluated based on the estimation of Neutral Detergent Fibre, cellulose, hemicellulose and Acid Detergent Lignin fractions and on the monosaccharide composition of the cell wall polysaccharides. The broad variation in yield and composition of biomass encountered in the panel confirms the potential of *L. mutabilis* as lignocellulosic feedstock and points out the value of this panel as breeding tool for the improvement of biomass quality. A genome-wide association study was conducted to identify single nucleotide polymorphisms (SNPs) associated with biomass quality, both across locations and per specific location. Scanning of 16,781 SNPs across the whole genome identified 51 unique SNPs for biomass quality, of which 5 were common either among traits or GWAS models. For each of the traits analysed, between 3 and 10 SNPs were detected explaining 2.7-15.9 % of the phenotypic variation. Underlying these loci, 28 genes were proposed as candidate genes for biomass quality. Important genes involved in cellulose and sucrose synthesis (*CESA4, SPP1,WRKY33, GONST2*), monolignol biosynthesis (*SKIP31, WAT1, CCR-SNL6*) and pectin degradation (*RAV1, PE*) were identified, and will require validation to confirm their value for application in *L. mutabilis* breeding.

Keywords: *Lupinus mutabilis*, biomass quality, SNP, molecular markers, agricultural residues, GWAS, CESA4

Introduction

The need to lower the dependency on non-renewable energy sources has become a global priority. Replacing petrochemical feedstock with sustainable biomass production for bioenergy and other biobased products represents a key step in this process. Bioenergy cropping as it has been done in the past years has led to monocultures of major crops exclusively for biofuel production (e.g., maize or sugarcane), attracting scepticism as energy cropping becomes a direct competitor with food production. However, the use of other biomass sources such as agricultural residues, discarded as "lignocellulosic waste" after harvest, can represent a more sustainable alternative. Due to their high availability, agricultural residues represent an important renewable source of cellulose and fermentable sugars. Moreover, these resources have a low energetic and economical costs, are recyclable, have a low carbon footprint and can contribute to simplified waste disposal and added-value production for farmers (Garcia *et al.*, 2016; Pennells *et al.*, 2021).

Different sources of lignocellulose vary in chemical composition, cell wall structure and degree of recalcitrance to deconstruction, which in turn make biomass more or less suitable for the production of different biobased products. In general, lignocellulosic biomass consists of 35%–55% cellulose, 25%–40% hemicellulose, and 15%–25% lignin (Kumar *et al.*, 2009), which are the main fractions of plant cell walls in monocots. Biomass quality is defined by the relative content of molecules of interest and by the recalcitrance of plant cell wall to deconstruction, which is key in the extraction process of target molecules (Pancaldi and Trindade 2020). Modern breeding technologies can be used to optimize crops to suit the industrial needs and achieve an efficient exploitation of agricultural residues, e.g., by selecting for favourable alterations in cell wall properties.

In the past decades important developments in research have paved the way for the breeding of plants for biofuel and other biobased products. The advancement in molecular marker technology and the genetic analysis of natural variation have made it possible to identify genomic regions that are associated with improved biomass quality, and to deploy these regions through breeding. Several studies have reported the identification of quantitative trait loci (QTL) linked to biomass quality for major feedstock crops such as maize (Méchin *et al*,. 2001; Torres *et al.*, 2013; Li *et al.*, 2016; López-Malvar *et al.*, 2021) or miscanthus (Gifford *et al.*, 2015; van der Weijde *et al.*, 2017), but also for agricultural residues such as rice and wheat straw (Malik *et al.*, 2019; Nguyen *et al.*, 2020). These QTL reflects mainly natural variation in lignin, cellulose and hemicellulose content and have aided the selection of genotypes with improved digestibility and/or content and functionality of specific cell wall components. Furthermore, the increasing use of genome wide association studies (GWAS) as an approach to dissect the genetic architecture of complex traits, has also contributed to pinpoint relevant genes for the improvement of biomass quality. This approach, in combination with the unravelling of cell wall structure and synthesis in model species, has resulted in a list of homolog genes that could be mined or manipulated to improve biomass quality through pathway engineering. Gene families

involved in the synthesis of cellulose, hemicellulose and lignin along with genes affecting cellulose crystallinity, hemicellulose substitutions and monolignol composition in lignin polymers have been targeted as they are the most relevant for biomass quality (Pancaldi and Trindade 2020).

The development of tailored multipurpose crops for both food and non-food applications, can enable the profitable processing of agricultural "waste" into added-value products. In doing so, multipurpose crops can become crucial in meeting biomass demands without competing for food production or increasing environmental pressure (particularly on soil and land). Examples of comprehensive use of lignocellulosic residues are the use of tomato leaves, stems and juice by-products for the realization of antibacterial fibre-based packaging (Tiekstra and Adriaanse 2014), the preparation of composites based on pea hulls and starch (Chen *et al.*, 2009) or the use of rice and wheat straw for the production of biofuels and other biobased chemicals (Garcia *et al.*, 2016; Tian *et al.*, 2018; Schnitzer *et al.*, 2014).

Within the frame of Horizon 2020, the project LIBBIO aimed at investigating the possibility of growing *Lupinus mutabilis* as a multipurpose crop on European marginal land and the use of the non-edible parts as a source of lignocellulosic feedstock for the biorefinery industry. *L. mutabilis* is originally an Andean legume, but its introduction to Europe could have many advantages as it represents a potentially superior alternative to the current sources of protein and oil. Its high quality grains are characterized by a protein and oil content similar to that of soybean and its ability to fix nitrogen and to adapt to low input farming in temperate conditions make this crop suitable for cultivation in marginal lands, also at higher latitudes (Chapter 2). Given the similar characteristics of Andean lupin and soy, there is evidence that *L. mutabilis* protein isolates can be as much valuable to the food and chemical industry as soy isolates (Carvajal-Larenas *et al.,* 2016). An advantage relatively to soy is that *L. mutabilis* can fit into sustainable cropping rotation schemes, requiring less water and pesticides, and adapting more easily to poor soils and higher latitudes (Gulisano *et al.*, 2019). In rotation with cereal crops, the ability to actively enrich the soil through nitrogen fixation and phosphates mobilization can reduce the need for nitrogenous fertilizers, provide valuable disease breaks and boost cereal yields (Gladstones, 1970). The recent evaluation of over 200 accessions of *L. mutabilis* across Europe has revealed a high biomass production in Central-North european conditions, particularly for indeterminate types of Andean lupin (Gulisano *et al.*, 2022). Uses of Andean lupin as green-fodder or silage have been investigated also in combination with other crops such as pea (Mikić *et al.*, 2013) or maize (Austria, LIBBIO), but they are discouraged due to the presence of bitter alkaloids. Clearly debittered seeds are more profitable for food uses or oil extraction, therefore a separate utilization of agricultural residues as biomass source for the biorefinery industry is desirable. To our knowledge, the use of lupin agricultural residues for the biorefinery industry has been investigated so far only in two minor species, *L. nootkatensis* and *L. rotundiflorus*. Acceptable yields of fermentable sugars have been reported in both cases, confirming the potential of lupin for biogas and bio-methane production (Radillo *et al.*, 2011; Kamm *et al.*, 2006). The aim of this work is to analyse the biomass quality of *L. mutabilis* and to identify QTL and candidate genes associated

with biomass quality. The natural variation present in a panel of 223 *L. mutabilis* accessions was exploited to characterize the variation in biomass yield and composition across two different cropping conditions in Europe.

Materials and Methods

Plant materials

A panel of 223 *L. mutabilis* accessions was used in this study. The panel included mostly landraces, varieties, and wild material from the centre of origin of the species, the Andean region, provided by the Instituto Nacional de Investigaciones Agropecuarias of Quito (INIAP, Ecuador). Additionally, some breeding material was provided by European breeding institutes, namely the Instituto Superior de Agronomia (ISA, Lisbon, Portugal) and the Julius Kühn-Institut (JKI, Quedlinburg, Germany). Overall, the panel included material originating from Ecuador (95 accessions), Peru' (65), Bolivia (15), Chile (1), Belarus (7) and of unknown origin (18) and lines resulting from the ISA (18) and JKI (4) breeding programs. The panel was tested in three field trials: in Portugal as a winter crop between November and May in 2019, in The Netherlands as a summer crop between April and October in 2019 and 2020. Each location had a randomized complete block design with three biological replicates per accession. The experimental units were plots including 20 plants (4 rows with 5 plants per row), at a distance of 30cm x 30cm. Phenotyping of morphological and phenological parameters was conducted on the 6 central plants of each plot , to gather data on biomass yield (fresh weight), dry matter (DM) content and yield of agricultural residues (total dry biomass yield – dry seed yield). Detailed information about field trial design and data collection are described in detail in Chapter 3.

Biomass quality analysis

Mature stem tissues were collected per single plot, by harvesting the central part of the main stem (~20cm) of the 6 plants phenotyped per plot. For each accession, three replicates were obtained per location. Stem tissues were successively dried at 50 °C until constant weight and ground to 1 mm. A total of 1941 samples were processed and scanned using a Foss DS2500 near-infrared spectrometer (FOSS, Co. LLC, Denmark) to obtain representative NIR spectra for each of the samples under analysis. High-throughput phenotyping of biomass quality was achieved by estimation of cell wall traits using multivariate prediction models based on Near-InfraRed Spectroscopy (NIRS), after a calibration curve for each trait was developed. A subset of 110 samples was selected for calibration based on the variation of the NIRS spectra, using WinISI 4.9 statistical software (FOSS), and then biochemically analysed to develop predictive models. Sample selection is based on Mahalanobis distance (H) values, which measure the distances in spectral data from the mean population and from each other, to select the most diverse grouping of spectra (Mahalanobis, 1936) . For the biochemical analysis, the van Soest acid detergent fiber

method was used to assess fibre composition by determination of NDF (Neutral Detergent Fibre), ADF (Acid Detergent Fibre) and ADL (Acid Detergent Lignin) in dry stem tissues (Soest et al., 1991). Further, biochemical analyses of monosaccharide composition of cell wall polysaccharides were performed to precisely assess the content of glucose and of major pectic components, not quantified by the van Soest method. Specifically, content of glucose, galacturonic acid, rhamnose, arabinose and galactose were measured as % of the alcohol insoluble residue (AIR) which is a good estimate of the cell wall content, according to the protocol developed by Petit et al. (2019). Calibration equations for each of the trait biochemically measured were fitted using partial least squares (PLS) and modified partial least square (MPLS) regression methods implemented in WinISI software. The coefficient of determination of cross validations (RSQ) ranged between 0.68 and 0.95 and the ratio performance deviation (RPD = standard deviation/standard error of cross validation) between 1.5 and 2 indicating a good predictive capacity for all traits. Monosaccharide components were estimated and directly predicted as % of AIR (% AIR) while predicted detergent fibre fractions were used to calculate the concentrations (in g/kg dm) of cell wall (NDF), cellulose (CEL, equals ADF-ADL), hemicellulosic polysaccharides (HEM, equals NDF – ADF) and acid detergent lignin (ADL) in stem dry matter.

Phenotypic data analysis

Descriptive statistics for each trait were calculated using basic functions in R software. An ANOVA model was performed on phenotypic traits means to determine significant effects of genotype (G), environment (E) and genotype-by-environment interaction (GxE). The analysis of the data was performed both per individual field trials and on all environments, using adjusted means across trials. In addition, a random effects model was used to calculate variance components and estimates of broad sense heritability across trials. Broad sense heritability was calculated across the three environments as in Renaud et al. (2014):

$$H^2 = V_G/(V_G + \frac{V_{GE}}{nE} + \frac{V_\varepsilon}{nE * nBlock})$$

where VG, VGE, V_ε represent respectively the estimated genetic, GxE and error variance components, while *nE* represents the number of environments and *nBlock* the number of blocks in each environment. Phenotypic (Pearson's R) correlations between traits were calculated by generating a correlation matrix using the packages *Himsc* and *corrplot*.

Genotyping

DNA extraction

Genomic DNA was isolated from young grinded *L. mutabilis* leaves (~20–400 mg, freeze dried material) using acetyl trimethyl ammonium bromide (CTAB) method (Doyle *et al.*, 1987) with

additional steps to remove proteins, polysaccharides, and RNA (as described in (Petit, Salentijn, Paulo, Thouminot*, et al.*, 2020)). DNA samples were controlled for quality and DNA concentration on agarose gel and by Qubit^TM Fluorometric quantitation to provide high quality genomic DNA for massive sequencing. Due to a certain degree of outcrossing behavior in *L. mutabilis*, each accession used in the GWAS panel might have a certain degree of genetic heterogeneity, despite being phenotypically homogeneous. Therefore, to cover all allelic variation within accessions, the genomic DNA of ten individual plants per accession was isolated and pooled, resulting in 223 samples for genotyping by restriction site associated DNA sequencing (RAD-seq).

RAD sequencing

The GWAS mapping panel was sequenced with RAD-seq to identify single nucleotide polymorphisms (SNPs) distributed over the genome to be used as molecular markers. 1 ug of high-quality genomic DNA (at a concentration ≥ 25 ng/μl) was digested using the restriction enzyme EcoRI. Then, RAD libraries with insert sizes of 300–550 bp, were prepared for each sample. The 223 samples were paired end sequenced on the HiSeq 4000 or Xten System (Illumina platform) to provide 1 Gbp genomic data per sample. RAD library preparation and sequencing were performed by Beijing Genomics Institute (BGI, Hong Kong)

RAD-Seq Data Analysis

Adaptors from the sequences were trimmed and low-quality reads were removed. Low quality reads comprised reads with > 50% of the bases Q ≤ 10, reads with unknown bases > 10%, reads that contain the sequence of the adaptors and reads that lacks a part of the multiplexing barcode, and could not be identified. The clean sequence reads of each sample were mapped to the *L. angustifolius* 'Tanjil' (LupAngTanjil_v1.0 refSeq GCF_001865875.1) genome reference (Hane *et al.* 2017) using Burrows–Wheeler Alignment Tool based on BWA- MEM algorithm. Picard-tools (v2.7.1) was used to sort the Sequence Alignment Map (SAM) files by coordinate and convert them to Binary Alignment Map (BAM) files. The average mapping rate was 76.9%, and the properly paired average 63.7%. Subsequently, BCFtools (v1.9) was used to call SNPs in each sample based on genotype likelihoods (specific BCFtools mpileup parameters: (d) max-depth =10000). In total, 3.747.406 putative SNPs were identified.

SNP Marker Selection

Samples for each accession consisted of pools of ten diploid plants to cover all allelic variation present in each accession. Each sample therefore harbors DNA from 20 alleles, represented by mostly expected two different nucleotides but occasionally three and rarely four different nucleotides (A, G, C, and T) at a position. For each polymorphic site, the allele frequencies were calculated per accession and in the GWAS panel.

Quality SNP marker selection was performed based on a 100% call rate of the SNPs in the 223 *L. mutabilis* accessions. Markers with a minor allele frequency below 5% and with a major allele frequency above 95% in the mapping panel were removed. Only biallelic markers were selected,

having a frequency sum of the two major alleles equal or above 95% and a sequencing depth >40x. After removal of SNPs not yet assigned to any chromosome, a total a set of 16,781 SNPs was selected for the genetic analysis. Each SNP was scored as the proportion of the major allele in the pooled sample of plants from the same accession. Quality SNP marker selection was performed in Python using a custom script. Finally, missing values were imputed using the average frequency of the concerned marker to minimize unwanted shifting of the data.

Estimation of Linkage Disequilibrium

Linkage Disequilibrium between SNPs across the genome was estimated on *L. mutabilis* GWAS panel following the approach of Vos et al. (2017). Short-range LD decay was based on the 90th percentile of Pearson allele frequencies correlation (r^2). The LD decay graph was drawn by fitting a smooth spline of r^2 over physical distances, measured as pairwise differences between marker position (Kbp), using the RQSMOOTH procedure in Genstat. Critical values of r2 as evidence of linkage, were set based on a fixed value of 0.1 (Remington *et al.*, 2001; Nordborg *et al.*, 2002; Robbins *et al.*, 2011). The intersection between the smooth spline line and the baselines (r2= 0.1) was used to define the distance at which LD decays.

Genome Wide Association study and analysis of population structure

A single trait GWAS was performed for each cell wall trait individually using the efficient mixed-model association approach (EMMA algorithm;(Kang *et al.*, 2008)) which was implemented in StatgenGWAS package v1.0.5 (van Rossum *et al.*, 2020). Based on the observation of the generated scatterplot of observed LOD-scores versus expected LOD-scores (Turner, 2018), the Van Raden kinship (2008) was selected as best correction in combination with the addition of the origin of accessions as a covariate. Finally, SNPs with a minor allele frequency below 0.02 (112 SNPs) were also excluded from the GWAS analysis. To characterize the genetics of *L. mutabilis* biomass quality across different environments and growing conditions, GWAS was performed across location (combining the three trials) and per location (PT, NL-Sc and NL-Wi) to identify also markers more specific for certain environments. To account for multiple testing and estimate the threshold for significant associations, two different corrections were used: a more conservative Bonferroni correction based on the number of independent markers at 5% significance (Li and Ji, 2005), leading to a threshold of 5.52 for $-\log_{10}(p)$, and one less stringent correction with a cutoff of $P=1\times10^{-4}$ for the detection of relevant associations.

Candidate gene identification

All candidate genes were selected based on the information contained in the NCBI *Lupinus Angustifolius* Annotation Release 100 for the genome assembly GCF_001865875.1 of LupAngTanjil_v1.0 (https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/3871/100/). The overlap of associated SNPs with annotated coding sequences in *L. angustifolius*, was determined using the SNPs positions available in the BED file. Based on the LD analysis, several genes in or near (within 80 Kbp up- and downstream of the lead SNP) the associated SNP were selected as

candidate genes. Special attention was given to genes with predicted functions related to the biosynthesis and modification of the cell wall.

Results

Biomass yields in different European cropping conditions

In this work, we assess the potential of *L. mutabilis* as biomass crop by investigating natural variation in biomass yield and quality in a panel of 223 accessions, across two different cropping conditions for Europe. Our results showed a high impact of cropping conditions on plant development and, as a consequence, on biomass yield (Figure 1). In winter Mediterranean conditions, on average were harvested 12.9 t/ha of fresh aboveground biomass with an average dry matter content of 56%, leading to a production of 4.9 t/ha of agricultural residues (dry weight). In summer-North European cropping conditions, 55.86 t/ha of fresh biomass were harvested in NL-Sc and 84.18 in NL-Wi, leading to respectively 14.7 and 20.9 t/ha of agricultural residues. The dry matter content of biomass at harvest was stable in North European trials (28.7 % and 26.8 % respectively in NL-Sc and NL-Wi), demonstrating that plants were harvested at the same developmental stage in both years.

Differences in biomass yield across the two trials in The Netherlands can be explained by the extreme heatwave hitting The Netherlands during summer 2019 and negatively affecting plant development. Conversely, the differences in biomass yield and dry matter content between Mediterranean and North-European cropping of *L. mutabilis* reflect the different development and growing stage of plants at harvest in these two conditions under study. In winter Mediterranean cropping, *L. mutabilis* ceased vegetative development soon after flowering as it encountered drought, reaching complete maturity at harvest with small and very dry plants. In summer North European conditions, a generally higher water availability prompted vegetative development by favoring a constant overlap between vegetative and reproductive phases until harvest, when plants had a bigger canopy and were still holding a higher water content. Nevertheless, the presence of a high range of variation in yield of biomass and agricultural residues within the panel under study, suggest that the selection of accessions for biomass production in Mediterranean and North-Central European environments has economic potential and can already provide 10-30 t/ha of agricultural residues for the biobased industry.

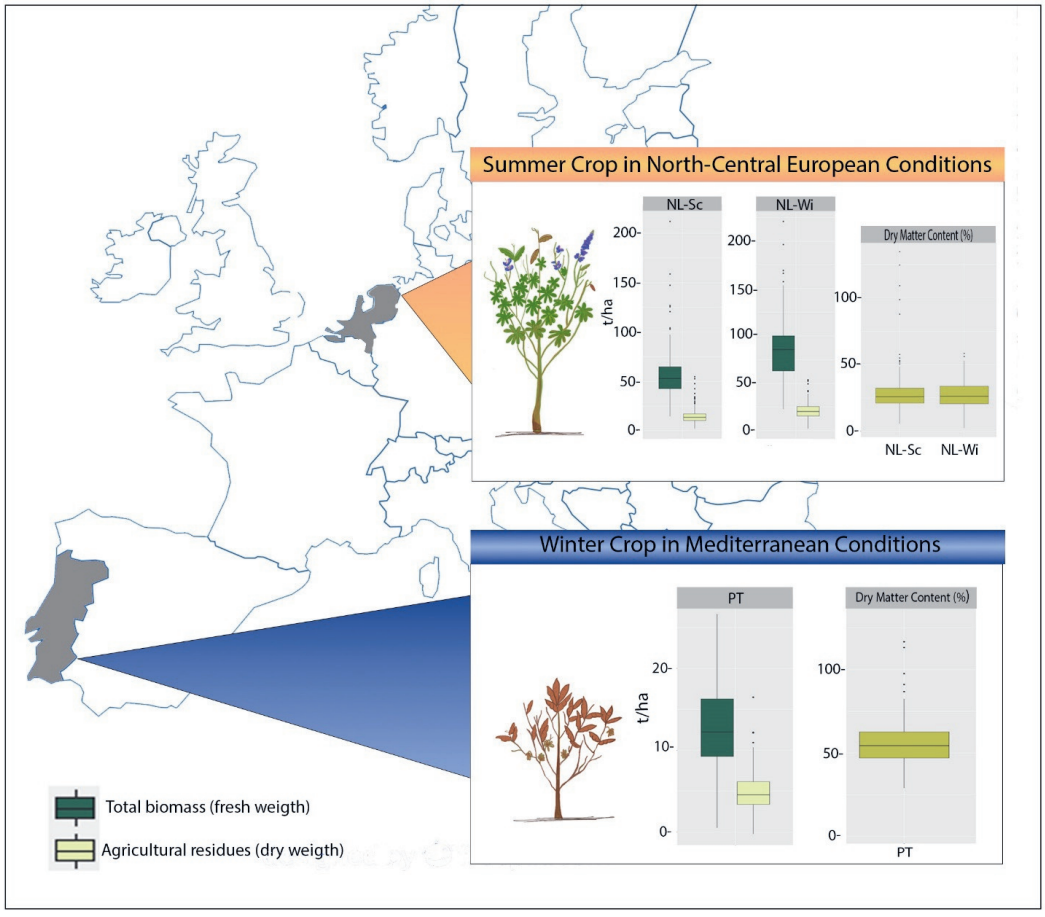


FIGURE 1| Potential yield of *L. mutabilis* as a biomass crop under two different cropping conditions in Europe. Fresh weight of total biomass (t/ha), dry weight of agricultural residues after seed harvest (i.e., stems and leaves, t/ha) and dry matter content (as % of total fresh weight) are reported for 223 *L. mutabilis* accessions growing in Portugal as winter crop and in The Netherlands as summer crop.

Variation and heritability of biomass quality traits in GWAS panel

This study represents the first evaluation of *L. mutabilis* biomass quality and shades light on the suitability of this crop as a source of biomass for the biobased industry. The presence of a significant range of variation for both the fiber fractions and the monosaccharide components of the cell wall measured across accessions (see Figure 2 and mean and ranges reported in Table 1) candidates this panel as a good reference for GWAS studies. Overall, cellulose and glucose represented respectively the higher fiber fraction (511.5±21.4 g/kg dm) and monosaccharide component (54±6.7 % AIR) of the cell wall, while hemicellulose and lignin were present in a similar ratio. We also report a high concentration of galacturonic acid (3.8 % AIR), constituting together with rhamnose the backbone of pectin in the cell wall. The remaining components, arabinose and galactose, are present in lower concentrations and can contribute to the formation of both pectic and hemicellulosic components of the cell wall.

Based on the phenotypic evaluation of the three different trials, the analysis of variance revealed a significant effect of genotype, environments and genotype by environment interaction (GxE) on each of the traits measured (P <0.01). Only for HEM GxE interaction effect was not significant. All the traits possessed a moderate to high heritability (ranging from 0.45-0.79) suggesting that variations of cell wall components are highly affected by genetic factors (see Table 1). When looking at mean values across environments (Figure 2), differences in fiber fractions content (NDF, CEL, HEM and ADL) across environments appeared not significant, while differences in monosaccharides compositions were significant (P<0.001) (depicted in Fig.2).

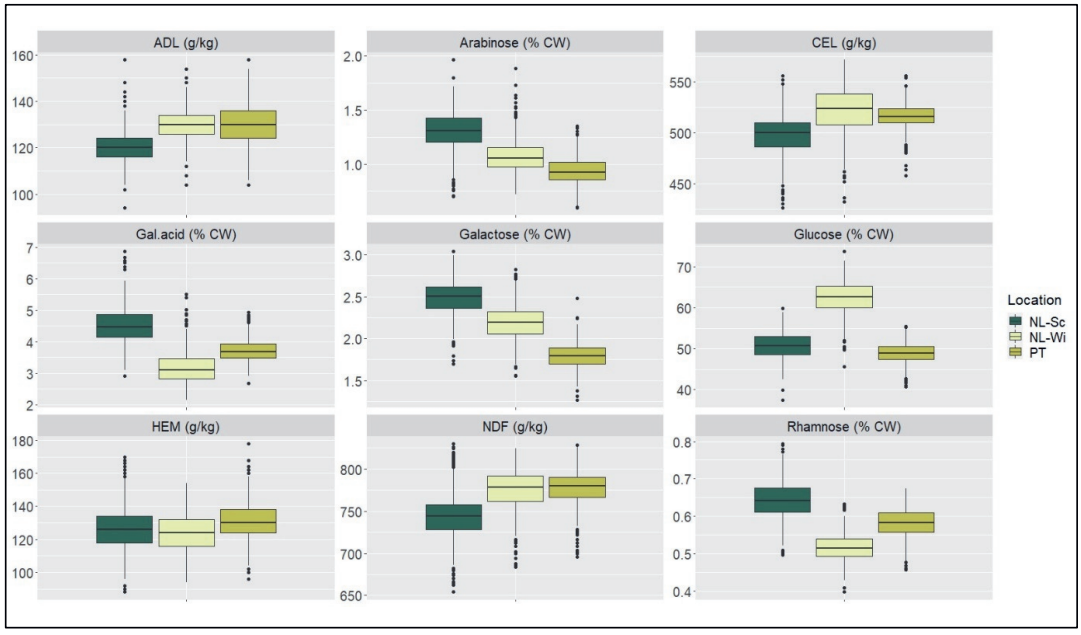


FIGURE 2| Predicted biomass quality of *L. mutabilis*. Differences in monosaccharide composition were all significant among the three trials (P<0.001), while differences in fiber fractions (NDF, CEL, HEM, ADL) were not significant.

The different level of accuracy in the biochemical assessment of fiber fractions and monosaccharide content might have contributed to this result, as shown also by the higher degree of variance associated to residuals for these traits (41.2-81.8%, Table 1). Nevertheless, our findings suggest that even when the yield of the different biomass fiber fractions would remain stable across locations, there are differences in the monosaccharide composition of these fractions. Furthermore, the differences in biomass quality traits observed between NL-Sc and NL-Wi, depicted in Fig. 2, suggest that even in similar locations meteorological extremes characterizing two different years can have a significant effect on biomass quality. Only galactose and arabinose appear as consistently higher in the cell wall of lignocellulosic biomass growing in North-European cropping conditions.

TABLE 1| Summary of statics for 12 biomass quality traits for 223 diverse *L. mutabilis* accessions (data averaged over the 3 trials). Mean and range represent the phenotypic variation span assessed for each trait; the distribution of total variance is shown as the percentage of variance components contributing to phenotypic variation; finally broad sense heritability (H^2) values indicate the extent to which a trait is genetically determined.

	Mean	Range	Variance components (%)					H^2
			Genotype	Environment	E: Block	GxE	Residual	
Rhamnose	0.58	0.40-0.79	6.5	69.0	3.8	2.3	18.5	0.70
Arabinose	1.11	0.60-1.97	5.7	54.4	9.6	3.7	26.6	0.57
Galactose	2.16	1.27-3.04	5.5	77.3	0.8	2.0	14.3	0.71
Glucose	53.77	37.48-73.69	1.8	82.6	0.8	1.3	13.5	0.48
Gal. acid	3.83	2.14-6.86	11.6	64.8	3.1	3.9	16.6	0.79
NDF g/kg	765.11	654-830	7.5	33.7	6.4	5.7	46.7	0.51
CEL g/kg	511.47	426-572	17.9	27.2	5.7	5.5	43.7	0.73
HEM g/kg	127.08	88-178	7.4	5.5	5.3	0.0	81.8	0.45
ADL g/kg	126.55	94-158	12.5	33.5	6.6	6.2	41.2	0.65

Relationships between biomass quality traits

A correlation analysis was performed between the nine traits analysed, including both fibre fractions and monosaccharide composition. To avoid relationships associated only to specific environments, correlation coefficients were calculated from the averaged phenotypic values across the three environments. As expected, positive correlations were detected between the fibre fractions of NDF and CEL ($r = 0.79$; $P < .001$) and NDF and HEM ($r = 0.44$; $P < .001$), as the NDF fraction comprises both hemicellulose and cellulose fractions. The positive correlation between ADL and HEM ($r = 0.28$; $P < .001$), components closely associated in the cell wall, and the negative correlation between ADL and CEL ($r = -0.38$; $P < .001$) are also in line with the known structure of plant cell wall. Furthermore, we detect a positive correlation between glucose and cellulose ($r = 0.73$; $P < .001$) but also between galactose and glucose ($r = 0.81$: $P < .001$) and consequently between galactose and cellulose ($r = 0.68$; $P < .001$). Instead, galacturonic acid, rhamnose and arabinose are highly correlated with each other ($r = 0.8$; $P < .001$) and share similar correlation patterns with the other traits, including negative correlations with NDF, CEL, HEM,

glucose and galactose and positive correlations with ADL. All the correlations between traits are depicted in Figure 3.

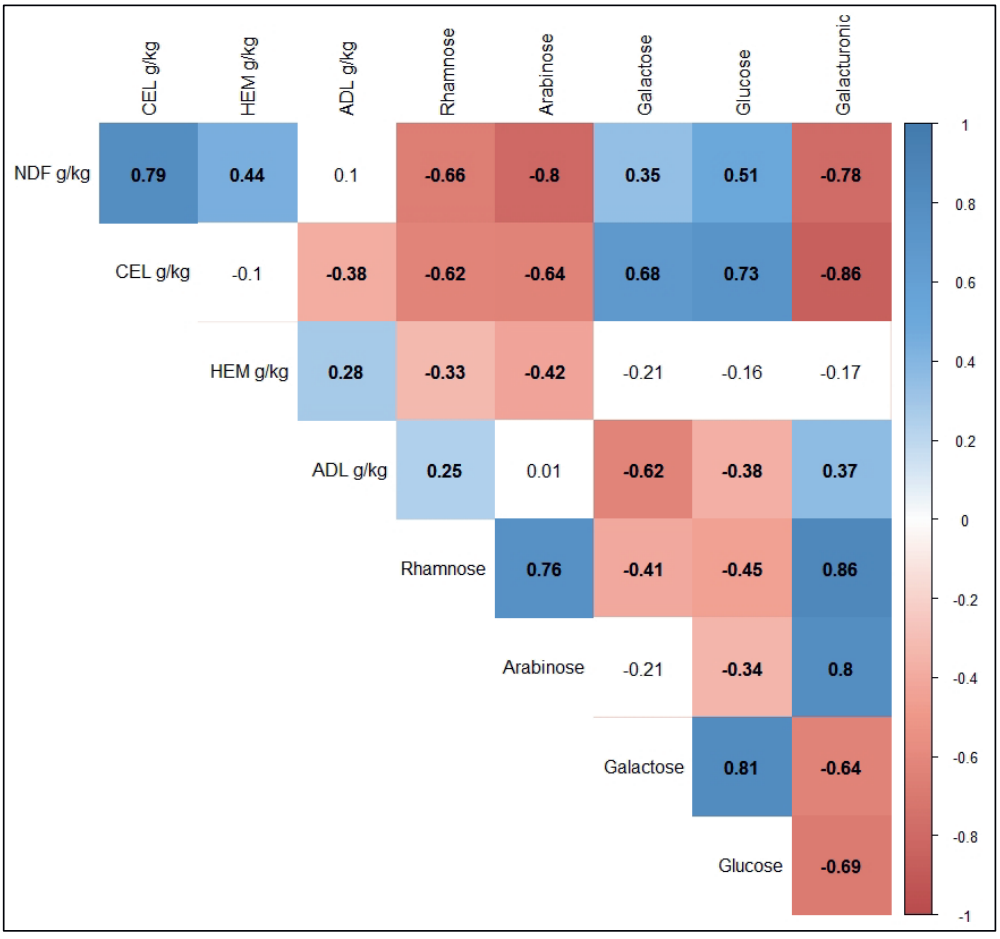


FIGURE 3| Phenotypic correlation between the 9 biomass quality traits analysed. Significant correlations were detected at a significance level of P value < 0.001 and blank cells represent not-significant correlations.

Genotyping of the GWAS panel

RAD-seq generated 2,053.375 million clean reads (~308 Gb) with on average of 9 million reads per sample and a read length of 150 bp. Following SNPs selection, a total of 187,468 SNPs markers were identified in the collection of 223 accessions of *L. mutabilis*. Out of these, 16,781 SNPs had a coverage > 40x and were physically mapped across the 20 chromosomes of *L. angustifolius* and were retained for all further analysis (Figure 4). An average of 839 SNPs were identified per chromosome, ranging from 479 SNPs on chromosome 14 to 1,477 SNPs on chromosome 11. The average marker density was approximately one marker every 28 Kbp across the 609-Mbp *L .angustifolius* genome (Hane *et al.*, 2017).

Linkage Disequilibrium and Population Structure analyses

Linkage Disequilibrium was estimated between all SNP markers over 223 accessions of *L. mutabilis*. Based on the intersection between the smooth spline line of the 90th percentile and the baseline for r^2 =0.1, LD was estimated to decay around 80 Kbp of distance (Suppl. Fig.1). A similar decay (LD = 77.45 Kbp) was recently estimated in a set of domesticated *L. angustifolius* accessions (Mousavi-Derazmahalleh, Nevado, *et al.*, 2018). The level of diversity and stratification present in the collection under study were also examined before performing the GWAS.

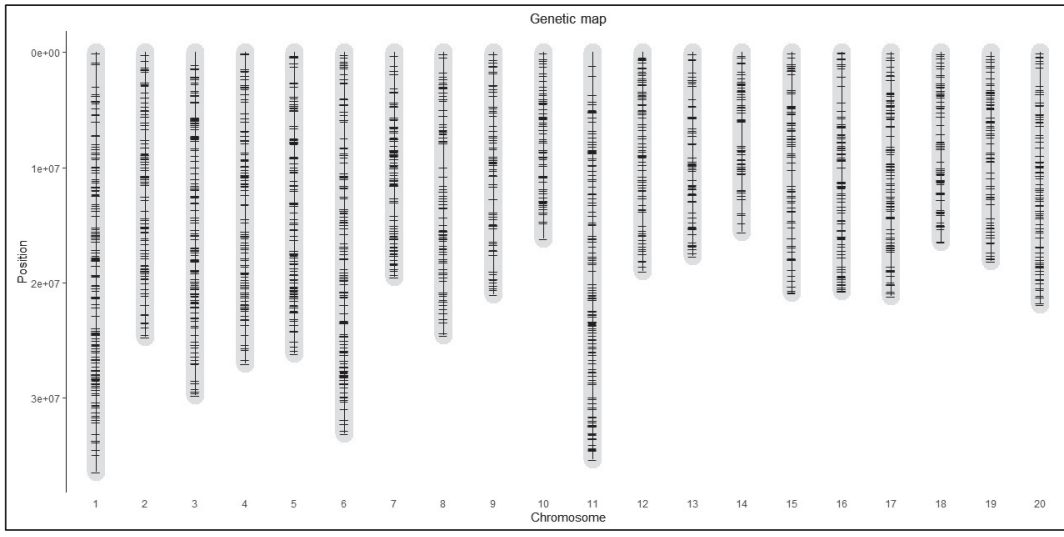


FIGURE 4| Genetic map showing the homogenous distribution of the 16,781 SNPs used in the GWAS analysis across the 20 chromosome of *L. angustifolius*.

Controlling for population structure is a standard procedure for GWAS but it becomes particularly relevant in this case, as our collection comprises genotypes from many different sources including both wild material and breeding lines. The Van Raden kinship analysis and the principal coordinate analysis of the distance matrix inferred from this kinship, revealed the presence of two subgroups in the collection (Fig. 5). This grouping confirms a division of the accessions based on their origin and type (wild material vs. breeding material). One group includes most of the wild material from Ecuador, Peru and Bolivia (see Fig. 5, red circle), while the seconds group consist mainly of breeding lines and elite material from Ecuador and Peru (see Fig. 5, blue circle). Even though PCoA1 and PCoA2 explained only 17.63% of the total genetic variance, this grouping corroborates previous findings. A similar separation pattern within this GWAS panel was already inferred from phenotypic observations (Gulisano *et al.*, 2022), and is now confirmed in a genomic relationship matrix based on genotype information for each individual marker (VanRaden, 2008).

GWAS of biomass quality in mature *L. mutabilis* stems

To dissect the genetic basis of natural variation in the cell wall components under study, a GWAS analysis was performed by fitting a linear mixed model with population structure and origin of the accessions. The analysis was firstly performed across-locations and then separately for each

of the locations, to identify both associated loci stable across locations and location-specific loci for the assessed traits. The quantile-quantile (QQ) plots of each trait confirmed that the associations were best controlled for population stratification by fitting both the Van Raden kinship matrix and the origin of accessions as covariate to the model. Manhattan plots and QQ plots generated by the GWAS analysis both across-location and per single locations are reported in Suppl. Fig. 2, 3, 4 and 5. At a significance level of $P < 1 \times 10^{-4}$ 17 SNPs were detected across-locations, of which 9 SNPs were identified for fibre fractions components and 8 SNPs for individual monosaccharide contents. For all the traits analysed, the phenotypic variance explained by each allele ranged from 2.7% to 11.8 %. In details, five SNPs were found in four loci significantly associated to NDF (corresponding to cellulose, hemicellulose and lignin), of which two were located on chromosome 8 and two on chromosome 9.

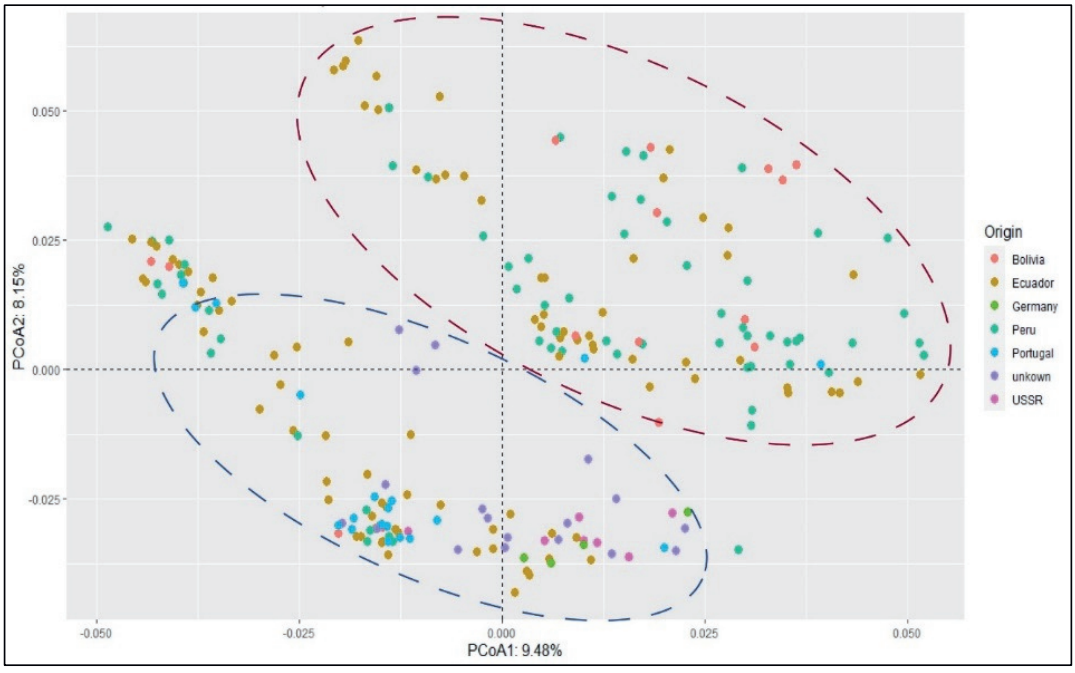


FIGURE 5|Principal Coordinates Analysis (PcoA) based on kinship matrix visualizes the estimated genetic structure of *L. mutabilis* collection under study and highlights separate grouping of wild material (red circle) and breeding lines (blue circle).

For cellulose (CEL) one significant SNP was detected on chromosome 14, while for hemicellulose (HEM) three SNPs were detected on chromosome 2, 3 and 11 respectively. One SNP was detected in association with ADL on chromosome 12. Regarding the monosaccharide content of cell wall polysaccharides, 1 SNP was associated to the content of arabinose (chr 20), 2 SNPs to galactose (chr 18), 2 to glucose (chr 6) and 3 to galacturonic acid (chr 8 and 20). No significant association with SNPs was detected for rhamnose across locations. In the location specific GWAS, 44 SNPs were found among all traits and locations. Of these, 20 SNPs were associated to fibre fractions and 22 to monosaccharide components. The number of markers detected and the explained variance of the model for the same trait generally differed across locations. Out of all detected associations, 5 SNPs were detected as common markers either

among traits or among the different GWAS models performed (i.e., across- locations and location-specific). SNP M7867 (chr 8, pos 23103579) -significant also at the higher Bonferroni threshold- was detected in association with NDF and galacturonic acid, while SNP M16431(chr 20, pos 11887436) was associated with arabinose and galacturonic acid both across-location and in the specific-location PT19. Similarly, SNP M7413 (chr 8) was associated with both arabinose and galacturonic acid and M10510 (chr 12) with NDF and galacturonic acid in NL-Sc, while M7865 (chr 8) was associated to galacturonic acid both across-locations and specifically in NL-Sc trial. Noticeably, among all traits and locations, we identify set of consecutive SNPs for related traits on different chromosomes, suggesting "hot spots" for hemicellulose and pectin related monosaccharides (such as arabinose, galactose, rhamnose and galacturonic acid) on chromosome 8 and 20 and for glucose on chromosome 6 (see Figure 6).

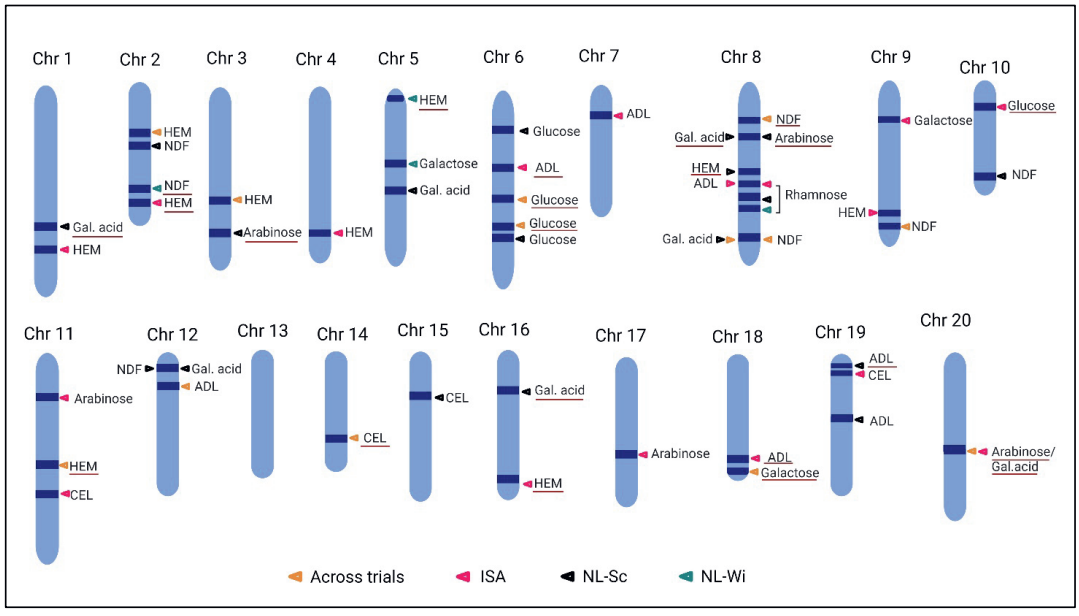


FIGURE 6| Schematic representation of the QTL found in association with *L. mutabilis* biomass quality traits and mapped on *L. angustifolius* genome. Colors of the arrow indicate the field trial locations where the respective SNPs were detected. The dark blue bars on the chromosomes indicate the confidence intervals for the QTL based on the linkage disequilibrium decay (± 80 Kbp flanking regions on either side of the associated SNP). The traits underlined in red are those for which relevant candidates genes with a role in cell wall composition were identified (Table 2).

Candidate genes for biomass quality in *L. mutabilis*

The reference genome assembly GCF_001865875.1 of LupAngTanjil_v1.0 on the NCBI genome database (https://www.ncbi.nlm.nih.gov/data-hub/gene/table/taxon/3871/) was used to annotate the genes that were near the significant SNPs. After excluding repeated loci due to co-localizing of SNPs for different traits and detection of consecutive SNPs leading to the same locus, a total of 68 genes were proposed as candidate genes across-locations for biomass quality

in *L. mutabilis*. Among those candidate genes, 17 encoded uncharacterized proteins, and the proteins encoded by the remaining genes include, transcription factors (ERF, bHLH, WRKY, bZIP type transcription factor families), enzymes involved in protein and lipid metabolism, stress response, photoperiod and flowering pathways, cell wall biosynthesis and other biological processes. Among the 68 genes, 9 genes were identified as associated to biomass quality across the three locations, in details to QTL for NDF, hemicellulose, cellulose, ADL and content of glucose and galacturonic acid (see Table 2).

In association with NDF, on chromosome 8, we find a gene encoding the transcription factor *RAV1*, a negative regulator of plant growth and development, and an *EXTENSIN-2-LIKE* gene responsible for the strengthening of primary cell wall. Instead, on chromosome 11, a gene encoding for WAT-1 related protein At2g37460-like (LOC109359103) and involved in the regulation of secondary cell wall thickness was identified in association with hemicellulose. Linked to the content of cellulose, we report the finding of an *EXPANSIN-A8-LIKE* gene (chr 14) causing loosening and extension of plant cell wall in Arabidopsis and rice, while in association with ADL we find *WRKY33* transcription factor (chr 12). *WRKY33* is a negative regulator of *CESA8,* a cellulose synthase terminal complex involved in the formation of secondary cell wall, which play a role in cell wall remodelling and stress response, in particular to drought (Wang *et al.*, 2013). Another transcription factor of the family WRKY, *WRKY15*, described as a negative regulator of cell wall lignification, was identified association with glucose content (chr 6). In association with glucose on chromosome 6 we also report the finding of *GONST2*, a GDP-mannose transporter associated with changes in cellulose content (Jing *et al.*, 2018). Lastly, two genes were associated to changes in galactose content on chromosome 18: *KINUA*, encoding a kinesin like protein KIN-UA functioning as a microtubule protein and influencing the structure and orientation of cellulose microfibrils (Nakamura *et al.,* 2021), and a gene encoding a LRR RECEPTOR-LIKE SERINE/THREONINE-PROTEIN KINASE probably involved in defence-related cell wall modifications/reinforcements. On chromosome 20, we identify a SUCROSE PHOSPHATASE (*SPP1*) involved in sucrose biosynthesis in relation with variation in galacturonic acid and arabinose content in the cell wall.

The location specific GWAS studies revealed the presence of new genes whose expression might be specific for certain environmental conditions (Table 2). In PT, eight genes associated to ADL, HEM and glucose content were identified. Lignin-related genes *PIP5K6* and *SKIP31* were identified respectively on chromosome 7 and 18. Furthermore, a *WAT1* (chr 18) regulating secondary cell wall thickness in stem fibres and *EXPA7* (chr 6) were also identified in association with ADL. Two genes on chromosome 2 were associated with hemicellulose, encoding a PECTINESTERASE-LIKE (*PE*) and a protein *STRUBBELIG-RECEPTOR FAMILY* 7-like (*SRF7*) involved respectively in pectin degradation and primary cell wall biosynthesis. Finally, a gene encoding a calmodulin binding protein with a role in biomass formation, protein IQ-DOMAIN 32-like *(IQD32),* was identified on chromosome 10 in a positive association with the content of glucose. In NL-Sc 14 genes were identified, associated with ADL, HEM and content,

arabinose, rhamnose, glucose and galacturonic acid. A CESA gene encoding a CELLULOSE SYNTHASE A CATALYTIC SUBUNIT 4 (*CESA4*) was identified for ADL on chromosome 19 together with *PER7,* encoding a PEROXIDASE 7-LIKE involved in lignin biosynthesis and degradation. Notably, on chromosome 8 we identify a variety of genes associated with hemicellulose and its components such arabinose and galacturonic acid. The SNP M7413 (chr 8, pos 7680541), co-localizing for arabinose and galacturonic acid, is contained in the gene *CRK10* which encodes a CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE 10 associated in *Arabidopsis* with dwarf genotypes and collapsed xylem vessels. In the same region we also identify genes encoding an EXTENSIN and a BETA-GLUCOSIDASE BOGH3B-LIKE involved in xyloglucan degradation. In addition, two more genes were associated with galacturonic acid: *EXPA4,* already identified across location on chromosome 16, and *FPP3* encoding a FILAMENT-LIKE PLANT PROTEIN 3 on chromosome 1, with a role in secondary cell wall deposition. Finally, two more genes associated with Arabinose were found on chromosome 3, encoding namely an INOSITOL-3-PHOSPHATE SYNTHASE-LIKE (*INO1*) involved in the synthesis of cell wall pectin and the transcription factor *RLM1,* which activates the transcription of genes involved in cell wall biosynthesis and maintenance.

In the last location, NL-Wi, the lowest number of associations was found and only three candidate genes were identified. On chromosome 2, linked to NDF, two genes encoded cell wall associated enzymes acting in the modification of cell wall via demethylesterification and de-esterification of cell wall pectin (a *pectinesterase-like* and a *pectinesterase 3-like*). Finally, a *cinnamoyl-CoA reductase-like SNL6* probably involved in lignin biosynthesis was identified on chromosome 5 for hemicellulose.

Discussion

The identification of QTL and putative candidate genes for *L. mutabilis* biomass quality has relevant implications for the breeding of this crop, when aiming at the utilization of its agriculture residues as biomass sources for the biorefinery industry. To our knowledge, this is the first genetic study of *L. mutabilis* both including a substantial number of accessions and assessing biomass quality in this species. This analysis provides the first evidence that *L. mutabilis* cultivation in Europe can generate a consistent amount of biomass for the European biorefinery industry and delivers the first genetic tools developed for targeted breeding of this crop.

Cultivation of *L. mutabilis* in Europe can produce between 27 and 54 t/ha of agricultural residues, respectively in Mediterranean and North- European conditions. Notably, we find that the biomass composition of these residues is comparable to the composition of other current sources of biomass, such as miscanthus and sorghum, and superior to the one of other lupin species (Hodgson *et al.,* 2010; Zhao *et al.,* 2009; Ţîţei 2020). We report cellulose yields in *L. mutabilis* up to 500 g/kg DM, which outperform average yield of common biomass feedstock such as miscanthus (400 g/kg DM,(Hodgson et al. 2010)) and sweet sorghum (180 g/kg DM,(Zhao *et al.*, 2009)). Glucose (54%) emerges as the main sources of fermentable sugars in *L. mutabilis*, consistently with previous results reported for the analysis of biomass quality in *L. rotundiflorus*

and *L. nootkatensis* (Radillo *et al.*, 2011; Kamm *et al.*, 2006). In these studies, hydrolysis of plant material yielded between 48 and 55 % of total fermentable carbohydrates, from glucose and xylose. Further studies should investigate the presence of xylose also in *L. mutabilis*. However, taking into account the similar cell wall content of glucose between those lupin species, the conversion of *L. mutabilis* biomass into fermentable sugars appears possible and promising. In particular, accessions LIB001, LIB027 and LIB067 emerge in this panel for their higher cellulose content across the different environments. LIB215, instead, stands out only in the Netherlands, indicating the better adaptation of this accession to North-European conditions.

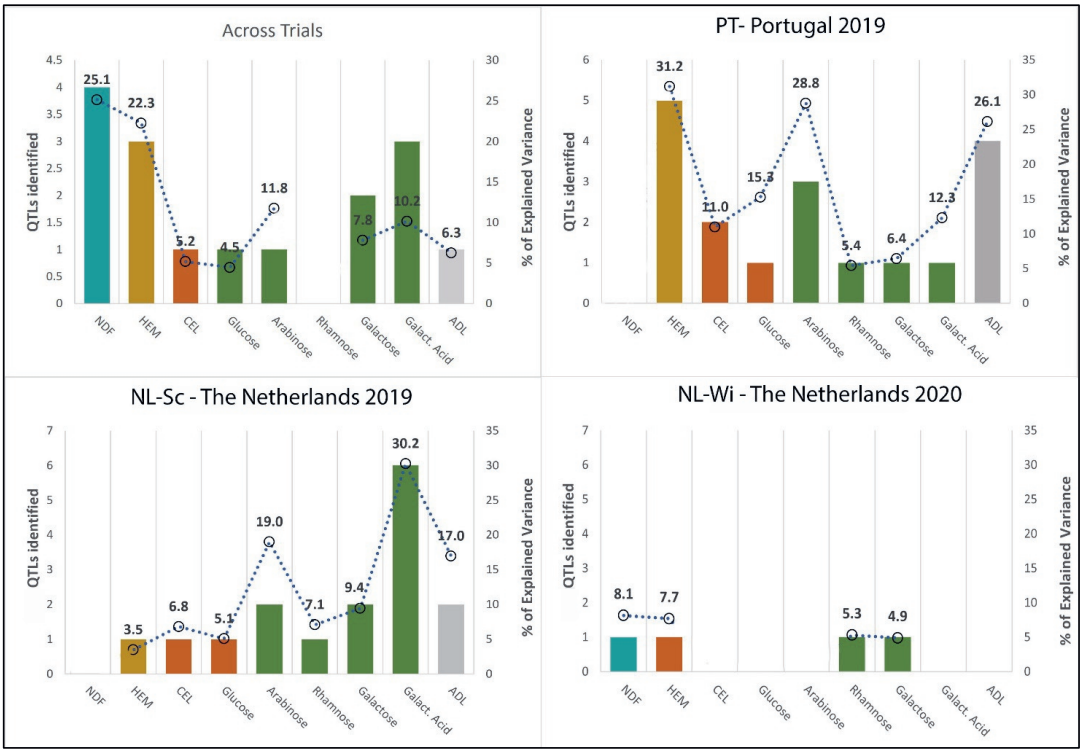


FIGURE 7| Survey of the QTL identified for biomass quality traits in a collection of 223 *L. mutabilis* accessions, across two different cropping conditions in Europe. Colored bars indicate total number of QTL identified per trait in our analyses. Circular points refer to the cumulative proportion of variance explained by all identified SNPs for a given trait.

TABLE 2| Putative candidate genes in QTL associated to biomass quality of *L. mutabilis* with a role in cell wall composition

Location	Trait	SNP name	Chr	SNP pos.	LOD	PVE	Gene ID	Gene name	Gene Description
All	NDF	M7437	8	7839123	4.59	5.71	LOC109354462		extensin-2-like
							LOC109353869	RAV1	AP2/ERF and B3 domain-containing transcription factor At1g51120-like
	HEM	M9798	11	21441300	4.08	4.73	LOC109359103	WAT1	WAT1-related protein At2g37460-like
	CEL	M11863	14	10434333	4.23	4.48	LOC109327136	EXPA8	expansin-A8-like
	ADL	M10606	12	3492591	4.28	6.28	LOC109361641	WRKY33	probable WRKY transcription factor 33
	Galactose	M15279	18	14096121	4.58	4.13	LOC109332916		probable LRR receptor-like serine/threonine-protein kinase At1g12460
		M15280	18	14096122	4.06	3.69	LOC109333139	KINUA	kinesin-like protein KIN-UA
	Glucose	M5759	6	19564231	4.42	4.09	LOC109350121	WRKY15	probable WRKY transcription factor 15
	Glucose	M5931	6	23471229	5.2	5.28	LOC109350274	GONST2	GDP-mannose transporter GONST2-like
	Ara/ Gal. acid	M16431	20	11887436	5.18/ 4.7	13.2 / 5	LOC109336456	SPP1	sucrose-phosphatase 1-like (SPP1)
PT	ADL	M5620	6	14743057	4.12	6.17	LOC109350828	EXPA7	expansin-A7-like
							LOC109352006	PIP5K6	phosphatidylinositol 4-phosphate 5-kinase 6-like
	ADL	M15249	18	13811328	4.25	6.02	LOC109332980	SKIP31	F-box protein SKIP31
							LOC109332164	WAT1	WAT1-related protein At4g19185-like
	HEM	M2124	2	19114587	4.39	7.75	LOC109337735	SRF7	protein STRUBBELIG-RECEPTOR FAMILY 7-like
							LOC109337703	PE	pectinesterase-like
	Glucose	M8427	10	4374335	7.26	15.26	LOC109358624	IQD32	protein IQ-DOMAIN 32-like
NL-Sc	ADL	M15721	19	8815933	4.35	6.86	LOC109334350	CESA4	cellulose synthase A catalytic subunit 4 [UDP-forming]
							LOC109354492		pyruvate kinase isozyme A, chloroplastic-like
							LOC109333967	PER7	peroxidase 7-like
	HEM	M7490	8	12066409	4.14	3.49	LOC109354794		beta-glucosidase BoGH3B-like
	Gal. acid/ Ara	M7413	8	7680541	4.18	5.49/ 11.99	LOC109354452	CRK10	cysteine-rich receptor-like protein kinase 10
							LOC109353935	EXT	extensin-like
							LOC109353651	NDR1	protein NDR1-like
	Gal. acid	M1040	1	25847112	4.53	2.87	LOC109358630	FPP3	filament-like plant protein 3
		M12931	16	6337087	4.71	10.36	LOC109330343	EXPA4	expansin-A4-like
	Arabinose	M3320	3	21772968	4.05	7.01	LOC109343654	INO1	inositol-3-phosphate synthase-like
							LOC109343658	RLM1	transcription factor RLM1-like
NL-Wi	NDF	M2112	2	18998440	4.19	8.15	LOC109337693	PE3	pectinesterase 3-like
							LOC109337703	PE	pectinesterase-like
	HEM	M4404	5	378011	4.12	7.68	LOC109347281	CCR-SNL6	cinnamoyl-CoA reductase-like SNL6

The broad variation in yield and composition of biomass encountered in the GWAS panel confirms the high potential of germplasm collections as important breeding resources for the improvement of biomass quality. Large ranges of variation in biomass yield across environments can be attributed to the effect of different environmental conditions on *L. mutabilis* development (Chapter 3). This is confirmed by the dissection of phenotypic variation for biomass components across trials which sees the effect of the environment as the largest on these traits (Table 1). The heritability values scored for these traits across environments (H^2= 0.45-0.79) are consistent with similar studies on other crops (Li *et al.* 2016; Petit *et al,*. 2020) and suggest a relevant effect of genetic factors on cell wall composition. GWAS analyses have proven to be a powerful tool to detect genetic components controlling quantitative traits. This tool has been used in many studies to approach complex traits such as cell wall composition, leading to the identification of QTL and candidate genes in important biomass crops (Slavov *et al*., 2013; Li *et al.* 2016; Nguyen *et al.* 2020) but never in lupin. In the present study, we successfully used GWAS to link variation in the yield of biomass fractions and in cell wall composition of *L. mutabilis* to putative candidate genes. In the absence of a genome sequence for *L. mutabilis*, the use of the closely related *L. angustifolius* genome proved suitable. Furthermore, the inclusion of population structure and of the origin of the accessions as covariate into the mixed model of the GWAS, ensured a good control over false positive associations, as shown by the Quantile-Quantile plots for the trait assessed (Suppl. Fig. 2,3,4 and 5).

A total of 51 SNPs were detected in association with biomass quality traits across all chromosomes, except for chromosomes 13 and 14 (Figure 6). Overall, between 3 and 10 SNPs were identified for all the traits evaluated (Figure 7). The phenotypic variation explained per trait by each locus ranged from 2.7-15.9 %, supporting the idea that the genetic basis of biomass quality traits is mainly controlled by several minor effect quantitative trait genes. Similarly, large number of QTL for biomass composition have been reported in miscanthus (van der Weijde *et al.*, 2017), maize (Torres *et al.* 2013; Li *et al.* 2016) and many other species. The identification of single SNPs in association with different traits contributes also to highlight the complexity of the cell wall and the high interconnection between different cell wall components. For instance, co-localization of SNPs for arabinose and galacturonic acid on chromosome 20 are supported by the high correlation (r=0.8) between these cell wall components. Furthermore, large correlations between cell wall components imply also that QTL affecting one trait can be identified in association with other strictly correlated traits. This is for example the case for the detection of a *CESA4* gene involved in cellulose biosynthesis in relation with lignin (ADL) content. Negative correlations between cellulose and lignin content have been reported in many studies, and even though to a lower extent, they are also reported in our study (r=-0.4). From our identification of SNPs associated to changes in cell wall composition emerges a list of putative candidate genes that could play a key role in the genetic improvement of *L. mutabilis*. It should, however, be noted that these results are based on the use of *L. angustifolius* reference genome. Therefore, future work is required to confirm the exact localization of the putative candidate genes in *L. mutabilis* genome, once available, and to carry out the functional studies required to confirm their roles in cell wall biosynthesis.

Potential candidate genes and implications for the genetic improvement of *L. mutabilis* biomass

Identifying the genes associated with cell wall components is crucial for improving biomass quality through pathway engineering. Among the candidate genes found in the present study, we discuss the presence of some of the most relevant genes involved in cellulose production and deposition, pectin degradation and regulation of monolignol biosynthesis (see Table 2). Changes in these processes have been shown to be key to the determination of biomass quality, as they can lead to improvements in biomass production and saccharification but also significantly reduce biomass recalcitrance.

A major feature of "ideal" biomass is a high cellulose content, leading to high yield of fermentable sugars. In most lignocellulosic crops, including monocots, the main bulk of cellulosic biomass derives from secondary cell wall (SCW) cellulose, which is synthesized by non-redundant cellulose synthase (CesA) proteins CesA4, CesA7 and CesA8. We report the presence of a *CESA4* gene on chromosome 19, in correspondence with a decrease in lignin content. As it would be logically expected, overexpression of SCW CesA4 has resulted in greater cellulose content and increased total polysaccharide content, e.g. in alfalfa and shrub willow (Yang et al. 2010; Serapiglia et al. 2012). However, attempts in other crops have resulted in decreased cellulose content and reduced growth suggesting that direct translation of an engineering strategy from a crop to another is not always straightforward (Brandon and Scheller 2020). We also report the presence of *WKRY33*, identified across-locations, which is a direct transcriptional repressor of *CESA8*. The single knockout of this gene has led to large increases in cell wall thickness, stem biomass density and above-ground biomass, suggesting targeting of this gene as a possible strategy to generate additional cell wall biomass in the stems of bioenergy crops without affecting the health and growth habit of the plant (Wang *et al.*, 2010; Wang *et al.*, 2013). Conversely, *WRKY15*, another transcription factor of the WRKY family acts as a negative regulator of cell wall lignification and was identified in association with cellulose. Changes in *WRKY15* expression could be more deleterious for plant development as *WRKY15* negatively controls the development of xylem vessels, and thus critical for plant survival (Ge *et al.*, 2020). Furthermore, it has also been proposed that changes in plasma membrane components can also affect CesAs activity and cellulose production. This might be the case for *GONST2*, a GDP-D-mannose transporter responsible for sphingolipids glycosylation and linked to cellulose content (Jing *et al.*, 2018). Another strategy to increase cellulose biosynthesis, which does not directly target CesAs, has been found by engineering the sucrose synthase pathway that is responsible for the production of UDP-Glucose, the sole substrate for the biosynthesis of cellulose. Previous studies have suggested that sucrose synthase (Susy) proteins interact directly with the CesA complex, channelling UDP-Glc directly to cellulose biosynthesis (Fujii *et al.*, 2010; Stein and Granot 2019). Accordingly, our analysis highlights also the presence of genes related to sucrose synthase in association with variation in polysaccharides content and composition. The identification of the SNP M16431 co-localizing for arabinose and galacturonic acid on chromosome 20, revealed the presence of a *SPP1* gene encoding a SUCROSE PHOSPHATASE 1 in this region, which catalyses the last step of sucrose synthesis where sucrose-6-phosphate (Suc6P) is dephosphorylated and released as sucrose. Overexpression of sucrose-phosphate synthase has

been suggested to improve biomass production and saccharification (De Vega *et al.* 2021; Falter and Voigt 2016), and could find similar application in *L. mutabilis*. Moreover, LRR RECEPTOR-LIKE SERINE/THREONINE-PROTEIN KINASES, like the one we identify on chromosome 18, can also have an active role in triggering sucrose synthase production at the phloem level, which stimulates cellulose synthases resulting in cellulose overproduction (Ghosh et al. 2013). Finally, several other genes can be important or even critical to cellulose biosynthesis even when they lack structural similarity to CesA proteins or glycosyltransferases of the sucrose pathway (Brandon and Scheller 2020). This might be the case for the protein IQ-DOMAIN 32-LIKE identified by this study in positive association with glucose content on chromosome 10. This calmodulin-binding protein has been identified as tightly co-expressed with secondary CeSA genes in Poplar and highly upregulated during phases of enhanced cellulose biosynthesis (Badmi *et al.,* 2018). The direct impact of this gene on cellulose biosynthesis is not clear yet, but stable knockdown of gene *PdIQD10* in Poplar has resulted in plants displaying enhanced growth and significant changes in cellulose content and crystallinity and cell wall sugar composition (Badmi et al. 2018).

Other than increasing cellulose content, decreasing biomass recalcitrance remains a key step for an efficient and economical biomass utilization. The recalcitrance of plant cell wall to deconstruction and release of sugars for fermentation is dependent on many factors, strongly interconnected and difficult to dissociate (Zoghlami and Paës 2019). Lignin represents the main hindrance to deconstruction, by creating a physical barrier in cross-link with hemicelluloses that hides polysaccharides to degrading enzymes (Zhao, *et al.*, 2012; Li *et al.*, 2016). Important genes involved in the lignin pathway emerged from our analysis of single environments. We find a F-BOX SKIP31 protein and a phosphatidylinositol-4-phosphate 5-kinase signal transduction gene (*PIP5K6*) previously reported in association with the biosynthesis of monolignol components (Hodgson-Kratky *et al.,* 2021; Dauwe *et al.*, 2007) and three more genes associated with total lignin content (WAT1, CCR-SNL6, and a peroxidases encoding gene). Several studies have demonstrated the role of F-box proteins in in the post-translational regulation of phenylpropanoids, key enzymes in the lignin biosynthesis pathway (Zhang *et al.*, 2013; Yu *et al.* 2019). In particular, the gene encoding for F-BOX SKIP31 protein was found to be upregulated (2.7 fold changes) in sugarcane genotypes characterized by a high syringyl/guaiacyl (S/G) monolignol ratios (Hodgson-Kratky et al. 2021). Shifts in the S/G ratio have resulted in increased ethanol yields in many studies (Fu *et al.*, 2011; Jung *et al.*, 2013; Ho-Yue-Kuang *et al.*, 2016), suggesting the modification of the monolignol ratio as a viable option for improving the efficiency of enzymatic hydrolysis of biomass. Moreover, such an approach has been shown to not have negative consequences on plant development (Franke *et al.,* 2000; Stewart *et al*,. 2009; Ho-Yue-Kuang *et al.*, 2016), while an overall reduction of lignin content has been generally associated with yield penalty (Masarin *et al.,* 2011; Benjamin *et al.*, 2013; Jung *et al.*, 2013). Regulation of *WAT1*, cinnamoyl-coa reductase (*CRR*) genes and peroxidases genes have resulted in strong alteration of lignin content and increased sugar extractability *(Ranocha et al.*, 2010; Jones, Ennos, and Turner, 2001; Ruel *et al.*, 2009; Goujon *et al.*, 2003; Bart *et al.*, 2010). Notably, we report the finding of *WAT1* in association with HEM across locations and with ADL in PT, on chromosome 11 and 18 respectively. These findings confirm the important role of *WAT1* in

secondary cell wall formation and suggest the presence of orthologs of this gene in *L. mutabilis* genome, likely characterized by different functions in different environments. These result points out this gene as an interesting candidate for a reduction of total lignin content in *L. mutabilis* that could potentially have restricted, if any, impact in plant growth and biomass yield.

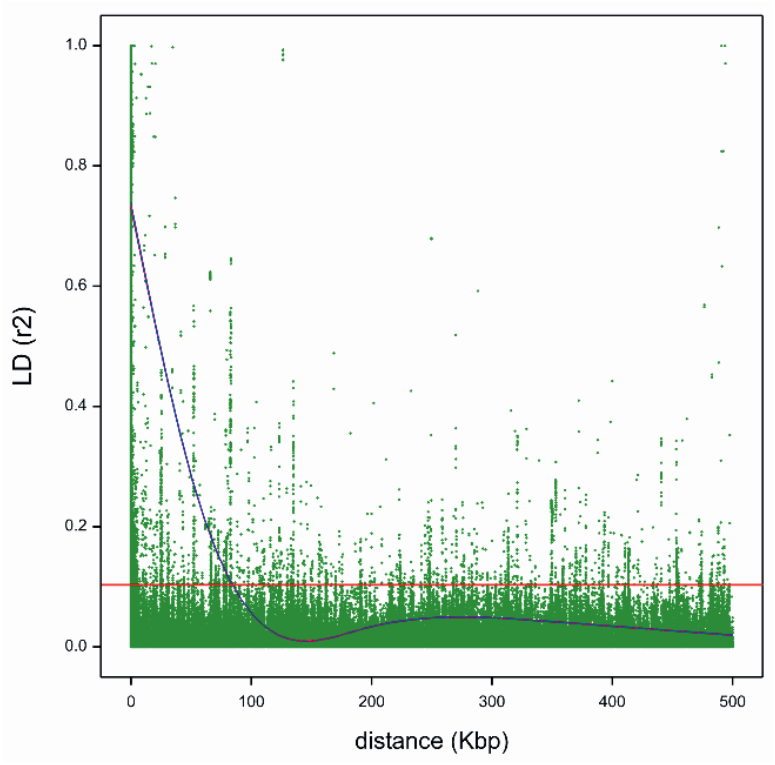
Pectin can also contribute to the recalcitrance of biomass by acting as a barrier to cellulolytic enzymes (Marriott *et al.*, 2016). Lionetti et al., (2010) have showed that reducing pectin content and increasing the amount of methyl-esterified pectin can increase biomass saccharification by 30-40%. We report the finding of a *RAV1* acting as a negative regulator of pectin degradation (across-locations) and of pectinesterases involved in the methyl-esterification of pectin (both in PT and NL-Wi). The recent identification of *RAV1* has proposed the presence of an important link between ethylene signalling in plant and pectin degradation, by showing that *RAV1* interacts with pectinases during early event of cell wall degradation (Tavares *et al.*, 2019). As suggested by the same authors, these results open up the possibilities to gain control on the process of cell wall degradation and facilitate cell wall hydrolysis for production of fermentable sugars (Tavares *et al.*, 2019). The pectinesterases found, act instead on the methyl-esterification of pectin. Similarly, inhibition of pectinesterases has been shown to reduce the amount of de-esterification of pectin in Arabidopsis, tobacco and wheat and to efficiently increase the efficiency of enzymatic saccharification, leading to higher release of glucose after enzymatic hydrolysis (Lionetti *et al.*, 2010). Finally, our analysis also identifies extensins and expansins responsible respectively for creating and loosening crosslinked networks in the cell wall, both across location and in specific-locations models.

Conclusion

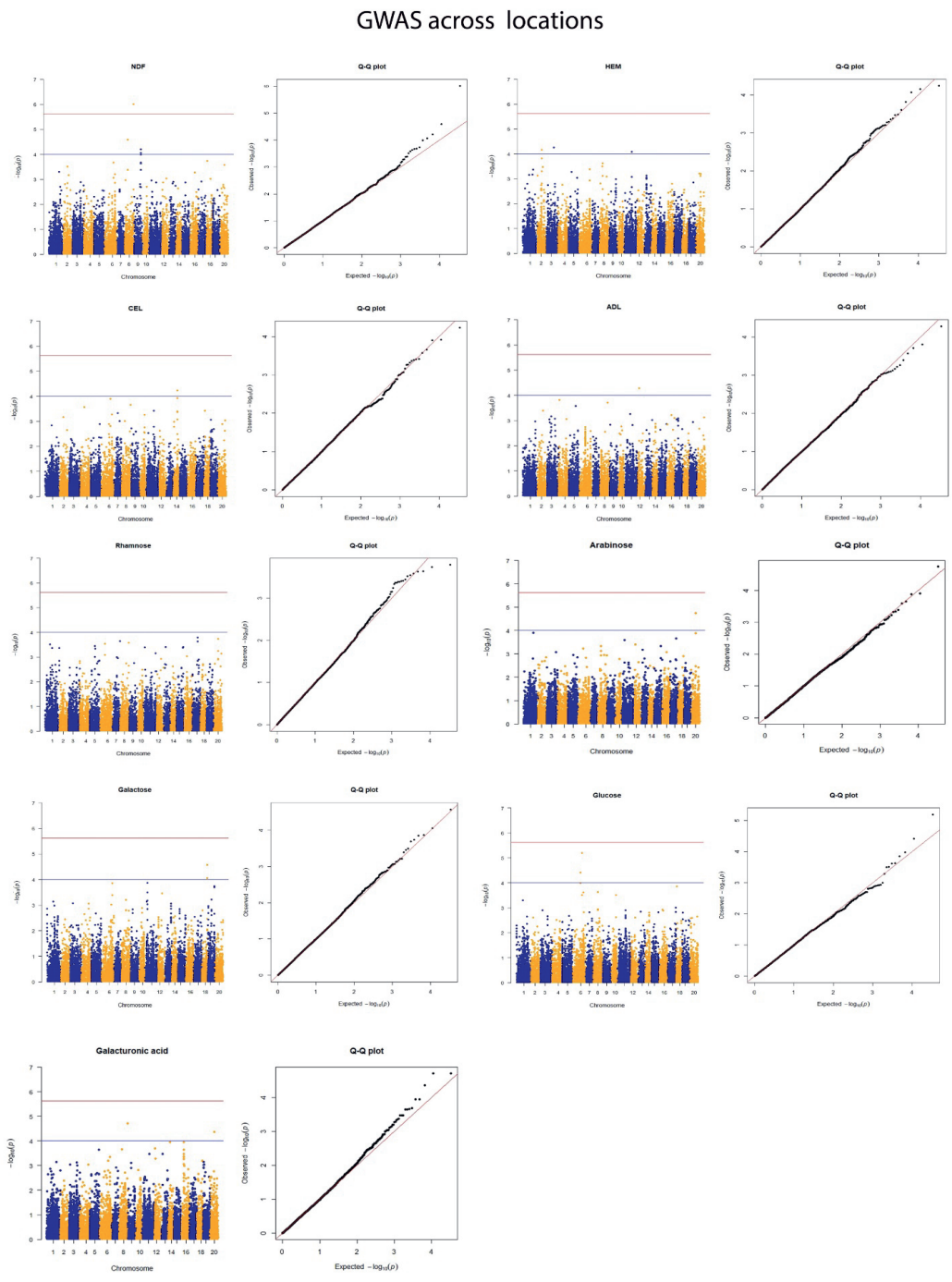
The present study is the first one to provide insights into *L. mutabilis* biomass quality and its genetic architecture. The results of this investigation reveal that *L. mutabilis* presents a valuable source of high-quality biomass, comparable or even superior to other the common biomass feedstock. The panel of 223 accessions selected for this study brings to light an unexploited wealth of heritable variation for cell wall compositional traits, which is key to the improvement of biomass quality. The GWAS analysis resulted in the identification of 51 SNPs for biomass quality and proposed relevant candidate genes associated with some of the QTL found. Genes involved in cellulose and sucrose synthesis could play an important role in tailoring cellulose content. Conversely, genes involved in the regulation of monolignol biosynthesis and in the degradation of pectin are good candidates for decreasing biomass recalcitrance and increasing the efficiency of enzymatic saccharification. After validation of the exact localization of QTL and candidate genes in *L. mutabilis* genome, the implementation of these findings in molecular breeding programs will unquestionably accelerate the development and introduction of *L. mutabilis* as a multipurpose crop for Europe. Furthermore, outstanding genotypes with high cellulose content have been identifies and they are valuable material to be included in breeding programs for biomass quality. Finally, the use of this *L. mutabilis* panel for the first genome-wide study on this species, has indicated that this collection of accessions includes not only an

extensive phenotypic variation, but also genetic. Hence, this panel can also be valuable for further genetic studies of *L. mutabilis* traits that are still poorly understood.

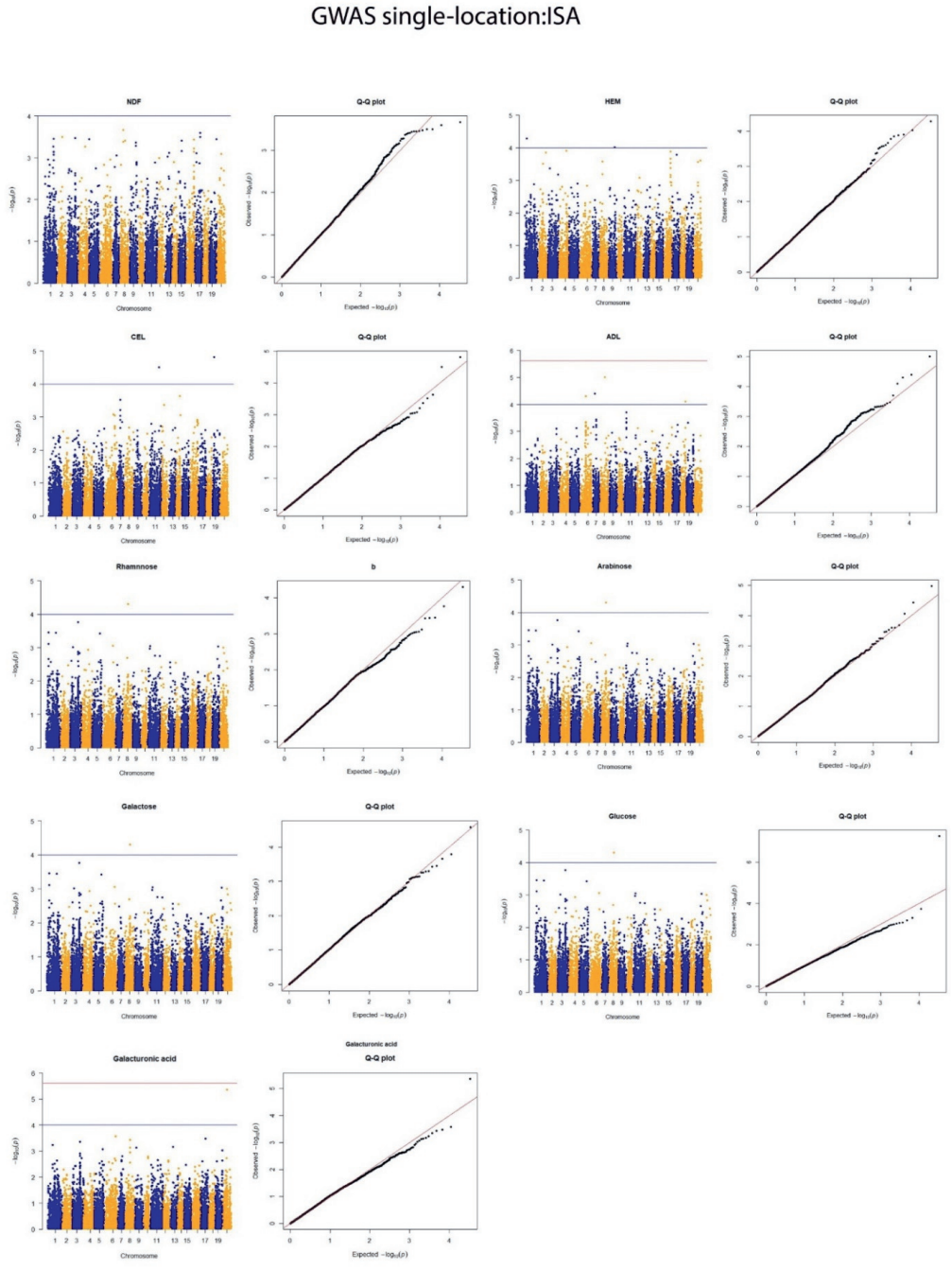
Supplementary Figures



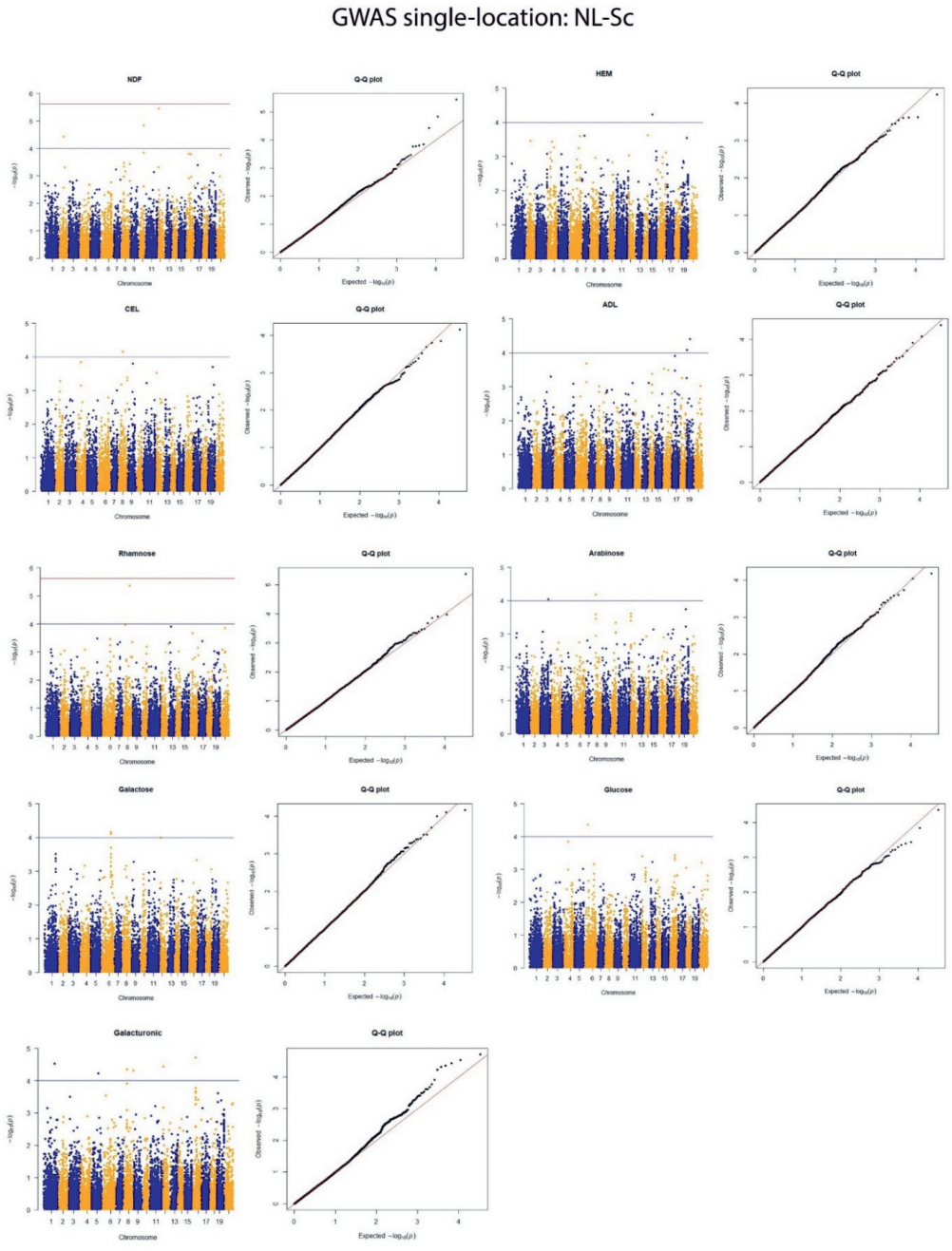
SUPPL. FIGURE 1| Plot of linkage disequilibrium. The blue line represents a smooth spline line fitted for a 90% quantile. The LD threshold of $r^2 = 0.1$ is indicated with a red line.



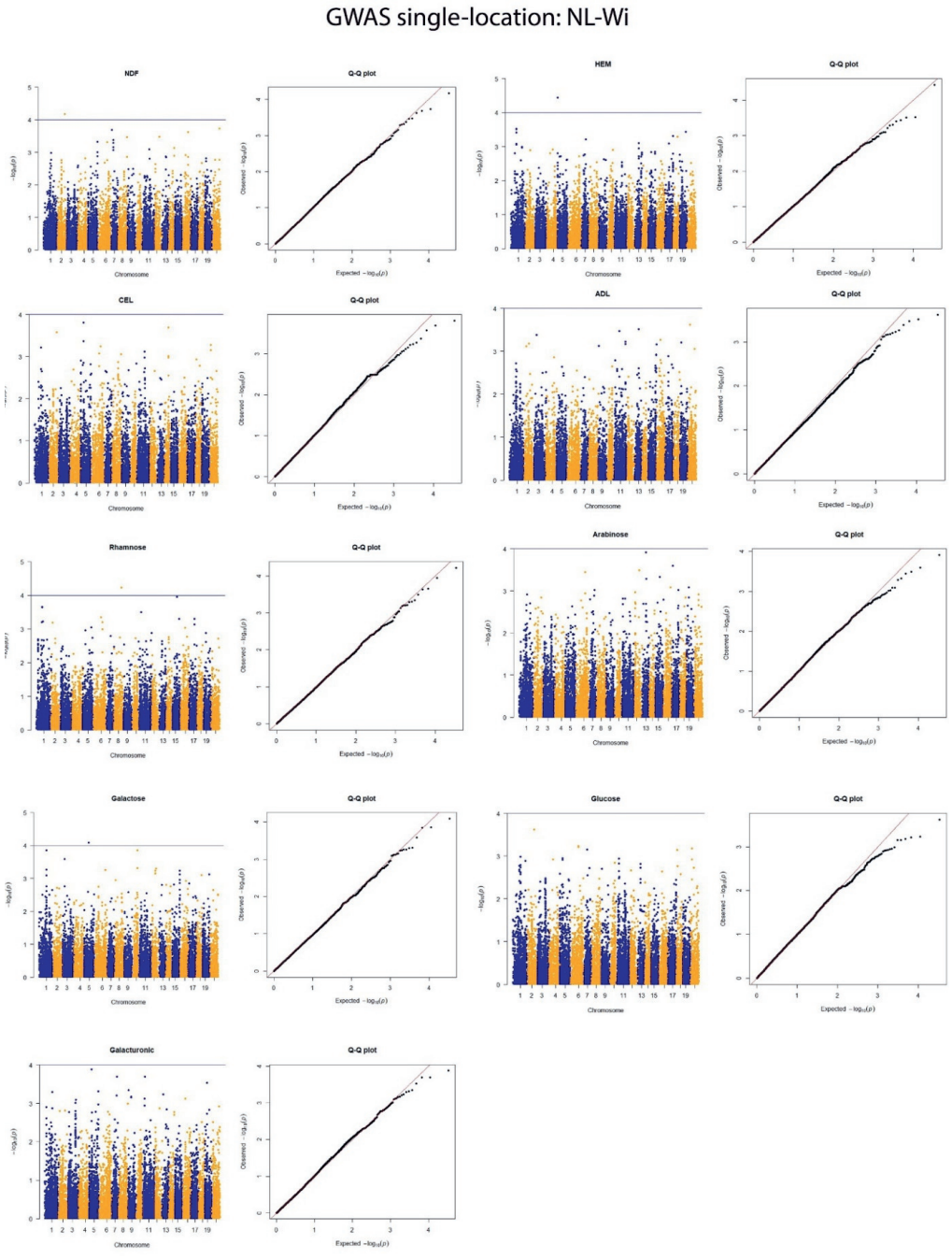
SUPPL. FIGURE 2| Manhattan plots and QQ plots resulting from GWAS analysis of *L. mutabilis* biomass quality traits both across all locations.



SUPPL. FIGURE 3| Manhattan plots and QQ plots resulting from GWAS analysis of *L. mutabilis* biomass quality traits per single location: ISA, Portugal.



SUPPL. FIGURE 4| Manhattan plots and QQ plots resulting from GWAS analysis of *L. mutabilis* biomass quality traits per single location: NL-SC, Scheemda, The Netherlands.



SUPPL. FIGURE 5| Manhattan plots and QQ plots resulting from GWAS analysis of *L. mutabilis* biomass quality traits per single location: NL-Wi, Winschoten, The Netherlands.

5

Genome-wide association study of agronomic traits in Andean lupin

Agata Gulisano[1], Antonio Lippolis[1], Eibertus N. van Loo[1], Maria-João Paulo[2], Luisa Trindade[1]

[1] Wageningen University & Research Plant Breeding, Wageningen University, Wageningen, Netherlands,
[2] Wageningen University and Research Biometris, Wageningen Research, Wageningen, Netherlands
[3] INIAP, Estación Experimental Santa Catalina, Quito, Ecuador,

Submitted.

Abstract

Establishing *Lupinus mutabilis* as a protein and oil crop requires the development of varieties with improved agronomic performance and adaptation to EU climates. Plant architecture, length of the growing cycle and yield are important breeding traits to select for, in order to ensure economic viability of *L. mutabilis* cultivation in European environments. However, due to a dearth of genetic and molecular investigations on this crop, it is currently unknown how these traits are regulated genetically in *L. mutabilis*. The aim of this study was to perform a genome wide association study (GWAS) to identify genetic markers, single nucleotide polymorphisms markers (SNPs), associated with these traits in *L. mutabilis*. A total of 16 669 SNPs were tested for association with 9 agronomic traits on a panel of 223 *L. mutabilis* accessions grown in four environments. Overall, a preponderant effect of the environment on the morphology and yield of *L. mutabilis* was observed, reflected by the detection of seven environment-specific QTL linked to vegetative yield, plant height, number of pods on the main stem and flowering time. The identified QTL revealed to have a major effect on the traits being able to explain 6 to 20 % of the phenotypic variation observed in these traits. Two stable QTL across environments were identified for flowering time on chromosome 8. The genomic regions surrounding these QTL harboured genes (FAF, GAMYB and LNK) involved in the regulation of major pathways influencing flowering and determinate/indeterminate growth habit, but also genes acting on these pathways in response to hormonal and environmental cues (GA30X1, BIM1, Dr1, HDA15, HAT3). These results are critical to initiate marker-assisted selection (MAS) and accelerate the development of *L. mutabilis* varieties adapted to specific cropping conditions in Europe, as well as to initiate functional studies on the role of genes regulating plant development and phenology in this species.

Keywords: *Lupinus mutabilis*, molecular markers, SNP, flowering time, GWAS, plant architecture

Introduction

Lupinus mutabilis, also known as "Andean lupin", is an endemic legume of the Andean region of South America. Firstly domesticated in the north of Peru (Atchison *et al.*, 2016), *L. mutabilis* has been traditionally cultivated in Ecuador, Peru and Bolivia for soil enrichment and as a food crop (Gross and Baer, 1981). Similarly to other Andean grains, in the last 500 years its cultivation has been marginalized and neglected due to the introduction of western pulses characterized by higher productivity. The scarce dissemination of its nutraceutical properties, and the presence of alkaloids providing a bitter taste to the seeds, have also partly contributed to this negligence (Chirinos-Arias *et al.*, 2015). Only in the past decades, the demand for alternative plant protein sources and the need to maximize agricultural land, by making use of marginal lands, has sparked renewed interest in this crop, not only in South America but also in Europe. *Lupinus mutabilis* seeds are characterized by a high content of protein and oil (44 % dm and 18 % dm, on average respectively), which exceeds that of any other lupin species and is comparable to soybean (Chapter 2). On top of that, *L. mutabilis* is adapted to low input farming in temperate climates and can effectively contribute to the improvement of poor soils by fixing nitrogen and mobilizing soil phosphates (Lambers *et al.*, 2013). In Europe and other temperate climatic regions, the combination of these features makes *L. mutabilis* a potentially superior alternative to current plant-based sources of protein and oil. Despite this, some challenges need to be addressed to expand its cultivation on large scale, as *L. mutabilis* remains to date an under studied crop. Pivotal to achieving this aim are breeding programs focused on guaranteeing economic viability and consumer acceptance of the crop.

In the past decades, numerous studies have investigated the nutritional profile and potential applications of *L. mutabilis* seeds in a wide range of end-use products, from protein, oil and food additives to cosmetics, medicines and bio-pesticides. In contrast, few studies have addressed the agronomic aspect of its cultivation. Research on the adaptation of *L. mutabilis* to European soil and climate conditions started only 30 years ago, when the first European project focusing on 16 selected lines was initiated (Caligari *et al.*, 2000). More recently, a second European project (LIBBIO) has expanded this investigation to 225 *L. mutabilis* accessions, by evaluating a wide panel of Andean germplasm both as a winter crop in Mediterranean conditions and as a summer crop in North-Central European conditions (Chapter 3). Both projects have pointed out the need of breeding for a better plant architecture, early maturity and yield stability (Caligari *et al.*, 2000; Chapter 3). These traits are highly interconnected. Plant architecture has been identified as the main factor limiting yield in European environments (Caligari *et al.*, 2000). This is to a great extent due to the indeterminate growth habit characterizing *L. mutabilis*, which leads to an overlap of vegetative and reproductive phases, hindering uniform maturation and delaying reproductive growth. The ability to prolong indefinitely vegetative growth after the onset of flowering, results in flowers and pods abortion in temperate climatic conditions during the long maturation period (R Galek *et al.*, 2007). Conversely, in Mediterranean environments, dry conditions at the end of

the cropping season can drastically affect biomass yield and pods set, decreasing considerably seed yield (A Hardy *et al.*, 1997). Early flowering genotypes can contribute to shorten the growing season and to escape terminal drought and heat stress (Chapter 3), but only the combination of determinacy and earliness can ultimately lead to higher and more stable seed yield.

Similar challenges have characterized the domestication of other legumes, including "old world" lupin species. Indeterminate growth habit is typical of many wild relatives of grain legumes (e.g. pea, soybean, common bean) and its switch to determinate growth can be considered as one of the most important traits of their domestication (Krylova *et al.*, 2020). The identification of genes responsible for growth habit has been pivotal for the development of determinate varieties, and has highlighted the interconnection of growth habit with stem length, flowering duration, yield, resistance to lodging and suitability to mechanized cultivation (Krylova *et al.*, 2020). Mutation of *Terminal Flower 1* (*TFL1*)-like gene controlling transition to flowering, has led to determinate growth habit in many legumes, including pea (Foucher *et al.*, 2003), faba bean (Avila *et al.*, 2006), common bean (Kwak *et al.*, 2008) and soybean (Z. Tian *et al.*, 2010). *TFL1* belongs to the small gene family *CENTRORADIALIS / TERMINAL FLOWER1 / SELF–PRUNING (CETS)*, controlling the developmental transition from indeterminate to determinate growth habit (Wickland *et al.*, 2015). Besides regulating flowering, genes of the *CETS* family are also involved in other processes, including stomatal opening or gibberellic and abscisic acid signaling pathways (Ando *et al.*, 2013; Xi *et al.*, 2010). Within the Lupin genus, determinate cultivars have been obtained in *L. albus, L. luteus,* and *L. angustifolius* through selection of spontaneous or induced mutants. The first determinate types in *L. mutabilis* have been obtained through induced mutation in Poland, distinguished by medium-tall stems without lateral branches, resistance to lodgings and early generative growth (R Galek *et al.*, 2007). Nevertheless, the molecular mechanisms underlying plant-architecture and other yield related traits remains unknown in *L. mutabilis,* due to the lack of genetic and molecular studies on this crop.

Given the importance of these agronomic traits, a large number of genomic regions (Quantitative Trait Loci, QTL) associated to plant-architecture related traits, flowering time and seed yield has been identified in the past decades for many grain legumes (Dargahi *et al.*, 2014; Yao *et al.*, 2015, Klein *et al.*, 2020; Ávila *et al.*, 2017). These studies have contributed to significantly increase knowledge on the genetic basis of these traits and to accelerate breeding of these crops. However, the majority of them has focused on identifying QTL in biparental populations, hence limiting genetic variation and mapping resolution (Gupta *et al.*, 2005). With the current development of sequencing and genotyping technologies at affordable cost, genome wide association studies (GWAS) have rapidly become a more common and powerful tool to investigate natural variation and to identify genomic regions underlying important agronomic traits. Moreover, GWAS approach allows to exploit higher phenotypic diversity than biparental mapping populations derived from targeted crosses, as well as a direct application of the results from research in breeding.

In this study, we evaluated a panel of 223 diverse accessions of *L. mutabilis* in the native Andean region, and over two cropping conditions in Europe, in order to identify Single Nucleotide polymorphisms (SNPs) underlying the variation in plant-architecture, flowering and yield related traits in this species. A GWAS approach was used to capture the natural diversity present in the panel, with the objective to develop genetic markers and highlight genomic regions (QTL) harboring causal candidate genes, which are critical for assisting and speeding up the breeding of *L. mutabilis*.

Material and Methods

Plant materials and field trials

The GWAS panel used in this study comprised 223 *L. mutabilis* accessions, 201 provided by the Instituto Nacional de Investigaciones Agropecuarias of Quito (INIAP, Ecuador), comprising mainly landraces, varieties and wild material collected across the Andean region, and 22 *L. mutabilis* lines developed in European breeding programs. The panel was evaluated in a total of four field trials in Ecuador and Europe during the growing seasons 2019 and 2020. The locations were chosen to represent an example of cultivation in the native environment as well as in Mediterranean (winter cycle) and North-Central European (summer cycle) climates and photoperiod regimes. The trial in Ecuador (EC, Cotopaxi 0° 55´ 35´´ S, 78° 40´ 07.4´´W) was sown in December 2019 and harvested in June 2020. The field layout was an alpha-lattice design with 3 replicates. Plants were arranged in plots with 5 rows (80 cm between rows), containing 40 plants spaced 20 cm. In Europe, the field trials were set up during winter (Nov 2019 – May 2020) for the Portuguese site (PT, Lisbon 38° 42´ 33.5´´ N, 9° 11´ 0.5´´W), and during two consecutive growing seasons from April to October in two locations in The Netherlands (NL-Sc 2019, Scheemda 53° 09´ 60´´ N, 6° 57´ 59.9´´E ; NL-Wi 2020, Winschoten 53° 10´ 11.´´ N, 7° 2´ 56.09´´E). A randomized complete block design (RCBD) with three replicates was adopted for the three European trials. The plots consisted of 20 plants sown at a distance of 30x30 cm. In all the locations, plants were cultivated under rain-fed conditions and without the aid of any fertilization, and following local cultivation practices.

Phenotyping of the GWAS panel

The phenotypic evaluation of the *L. mutabilis* GWAS panel across the four environments included the scoring of quantitative traits related to plant morphology, phenology and agronomic performance (Chapter 3). The phenotypes of interest were: flowering time, plant height, number of branching orders, vegetative yield, number of pods and seeds produced on the main stem, total number of pods and seeds produced on the overall plants, and 100 seeds weight. The traits were scored as described in our previous study (Chapter 3). Briefly, flowering time was scored as

number of days from sowing until 50% of the plants in a plot had started flowering. Due to unforeseen circumstances, longer intervals in scoring of flowering in NL19 resulted in suboptimal phenotyping, as also shown by the low heritability estimate for flowering in this trial (H^2=0.24; Chapter 3), hence these data were excluded from this study. Instead, in Ecuador, due to the impossibility of collecting data during Covid restrictions, scoring of flowering time and vegetative yield was not possible. Phenotyping was conducted on the six central plant of the plots (five in Ecuador). At harvest, height of the main stem (cm) and number of branching orders were scored (0 = main stem only, 1 = main stem and first branching order, etc.). Number of pods on the main stem (Pods MS) and total number of pods (Pods T) was also recorded. After harvesting, seeds were air-dried and counted separately on the main stem (Seeds MS) and on the total plant (Seeds T). Vegetative yield (dw, g/plant) was estimated as the difference between the total amount of biomass harvested (dw) and the seed yield per plant.

Statistical analysis

Phenotypic data were analyzed using the SpATS mixed model approach, implemented in StatgenSTA R package (v1.0.8), to correct for spatial gradients in the field by adopting a 2-dimensional smoothing with P-splines (Rodriguez-Alvarez *et al.*, 2018). The best linear unbiased estimations (BLUEs) of genotypic means were obtained from this model and then used for the rest of the analyses. Across trials, a random effects model was fitted using *lme4* R package to calculate variance components of genotype (G), environment (E) and genotype by environment interaction (GxE), and estimates of broad sense heritability. Broad sense heritability was calculated across the three environments as in (Renaud *et al.*, 2014):

$$H^2 = V_G/(V_G + \frac{V_{GxE}}{nE} + \frac{V_\varepsilon}{nE * nBlock})$$

where VG, VGEI, V_ε represent respectively the estimated genetic, GxE and error variance components, while *nE* represents the number of environments and *nBlock* the number of blocks in each environment. Pearson's correlations between BLUEs genotypic values in each trial were estimated and plotted using the R package *corrplot* (v0.92).

Genotyping and SNP development

Reduced representation sequencing and single nucleotide polymorphism (SNP) typing was performed as previously described in Chapter 4. Briefly, genomic DNA was isolated from young grinded *L. mutabilis* leaves (~20–400 mg, freeze dried material) using acetyl trimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987) as described in Petit et al. (2020). To cover some degree of genetic heterogeneity expected in the accessions, DNA was extracted from a pool of 10 individuals/accession. Restriction site-associated DNA sequencing (RAD-seq) was used to identify SNPs distributed over the genome, by digesting 1 ug of high quality genomic DNA (at a

concentration ≥ 25 ng/µl) using the restriction enzyme EcoRI. RAD library preparation and sequencing were performed by Beijing Genomics Institute (BGI, Hong Kong). The Burrows–Wheeler Alignment Tool based on BWA- MEM algorithm were used to map the clean sequence reads to the *L. angustifolius* 'Tanjil' (LupAngTanjil_v1.0 refSeq GCF_001865875.1) genome reference (Hane *et al.*, 2017). The average mapping rate was 76.9%, and the properly paired average 63.7%. BCFtools (v1.9) was used to call SNPs in each sample based on genotype likelihoods. At each locus, SNPs were called as percentage of the reference allele present on the total number of reads generated. After SNPs quality control and removal of SNPs not assigned to any chromosome, a total set of 16,781 biallelic SNPs was selected for the genetic analysis.

Genome-wide association studies

Single trait GWAS analysis was conducted separately for each environment, using 16,669 polymorphic markers after filtration to remove SNPs with minor allele frequencies (MAF) < 0.02. The GWAS model was based on a linear mixed model for association mapping as implemented in StatgenGWAS package v1.0.5 (van Rossum *et al.*, 2020). Single trait GWAS in StatgenGWAS follows the approach of Kang et al. (2008), by performing a two steps procedure. Firstly, an 'empty' model without any SNP effect is fitted in order to obtain REML-estimates of the genetic and residual variance components, computed using the Efficient Mixed Model Association (EMMA) algorithm (Kang, Zaitlen*, et al.*, 2008). Secondly, the single SNP-effect of interest is tested by using generalized least-squares (GLS) and F-test, obtaining the effect-size and P-values for all SNPs. Population structure and individuals relatedness were taken into account by fitting a Van Raden kinship matrix and adding origin of the accessions as corrections. The adequateness of genetic relatedness correction was assessed by evaluating genomic inflation factors. The Bonferroni correction was used to correct for multiple testing, thus obtaining a threshold of 5.52 for $-\log_{10}$ (p) that was used to state statistically significant SNPs. Manhattan plots were visualized using *qqman* package in R (Turner, 2018). Linkage disequilibrium for this panel was already estimated in Chapter 4, following the approach of (Vos *et al.*, 2017). LD was estimated to decay around 80 kbp of distance.

Candidate genes identification

The size of the genomic regions investigated to identify putative candidate genes, controlling the traits under study, was defined by the extend of the average Linkage Disequilibrium across the genome. Starting from the position of the detected significant SNPs, candidate genes were proposed whether harbored in a maximum physical distance, upstream and downstream, of 80 kbp. All candidate genes were selected based on the information contained in the NCBI *Lupinus angustifolius* Annotation Release 100 for the genome assembly GCF_001865875.1 of LupAngTanjil_v1.0 (https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/3871/100/). Special attention

was given to genes with predicted functions related to the flowering time and regulation of plant growth and development, based on gene description and relevant literature.

Results

Phenotypic analysis

Nine traits relevant for breeding of *L. mutabilis* accessions adapted to different cropping conditions in Europe and Ecuador were investigated using a panel of 223 accessions. Extensive phenotypic variations were observed for all traits, namely flowering time, plant height, number of branching orders, vegetative yield, number of pods and seeds on the main stem, total number of pods and seeds and seed weight (depicted in Figure 1). Remarkable variation in flowering time was observed between the two European cropping conditions. In particular, flowering occurred at an earlier time after sowing in the Netherland than Portugal, with an average of 81 and 110 days for NL-Wi and PT, respectively. Relevant phenotypic diversity in time to flowering was observed within the panel, as highlighted by coefficients of variation (CV %) of 6% in Portugal and 8% in the Netherlands 2020. The overall plant height across different locations ranged from 24 cm to and 248 cm, indicating that *L. mutabilis* small genotypes can be 10 times smaller than tall genotypes as combination of environmental and genetic effects. As a general trend, plants were shorter on average in PT (60 cm) and EC (73 cm) compared to NL-Sc (90 cm) and NL-Wi (92 cm), and characterized on average by a first branching order only. Contrarily, in NL-Sc and NL-Wi the main stem reached an average height of 90 and 92 cm respectively, with two/three branching orders produced. Plant height had high phenotypic variability within the 223 accessions, as pointed by the CV values of 17%, 18%, 20%, 13% in Ecuador, Portugal, NL-Sc and NL-Wi respectively. Vegetative yield reflected the development of *L. mutabilis* in different environments. In Portugal, the average vegetative yield recorded (45.8 g/plant) was between three and four times lower than the average yield of NL-Sc (133.6 g/plant) and NL-Wi (174.9 g/plant). Concerning grain yield components, the average number of pods and seeds produced by the total plant, as well as by the main stem only, was higher in winter-Mediterranean conditions, followed by the native environment (EC) and summer-North European environments. The number of seeds harvested per plant was on average 42 in EC (CV 47 %), 95 in PT (CV 19 %), 26 in NL-Sc (CV 49 %) and 42 (CV 53 %) in NL-Wi. The number of pods and seeds produced on the main stem were respectively 9 (CV 20 %) and 25 (CV 28 %) in EC, 13 (CV 23 %) and 42 (CV 21 %) in PT, 3 (CV 89 %) and 14 (CV 47 %) in NL-Sc and 9 (CV 27 %) and 20 (CV 37 %) in NL-Wi. Conversely, the yield of biomass agricultural residues was lowest in PT (45.8 g/plant) and highest in NL-Wi (174.9 g/plant). An exhaustive analysis of the phenotypic traits under study was conducted in a previous study of diversity and agronomic adaptation of this collection where the mean values, range of variation and Pearson's correlation between plant architecture and yield-related traits are reported (Chapter 3).

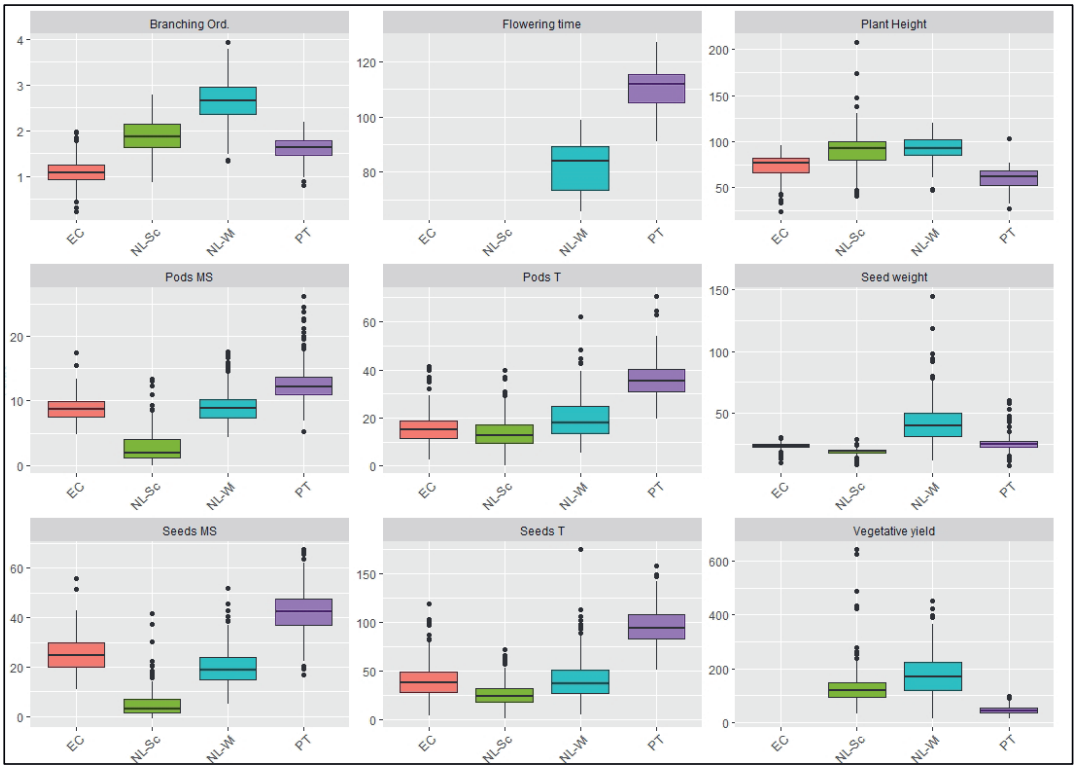


FIGURE 1| Boxplots showing the variation of the plant architecture and yield related traits (BLUES data) under study in a collection of 223 *L. mutabilis* accessions. The upper and lower end of the boxes indicate respectively the 75th and 25th percentiles, while the line in the middle represents the 50th percentile (median). The whiskers represent the highest and lowest values, while circles represent outliers outside the 5th-95th percentile interval.

Heritability estimates and identification of relevant breeding traits

In the four separate locations, estimates of heritability were high for flowering time (0.87-0.93) and plant height (0.5-0.8). Moderate to high heritability was also observed for number of pods on the main stem (0.4-0.8), number of seeds on the main stem (0.4-0.8), total number of pods (0.2-0.6), total number of seeds (0.2- 0.7), 100 seeds weight (0.1-0.67), and moderate to low value for vegetative yield (0.1-0.5) and branching order (0.1-0.4). When looking at the data across locations (averaged data), the analysis of variance across trials indicated that genotype (G) and environment (E) had a significant effect on all the traits analyzed ($P < 0.001$). The effect of genotype by environment interaction (GxE) was also highly significant ($P < 0.001$) for the majority of the traits, but had no significant effect on vegetative yield. As shown in Table 1, heritability values across locations were higher for plant height (0.82) and flowering time (0.69), two traits well known for being under the influence of a high quantitative genetic component in

many crops, as well as for the number of pods and seeds produced on the main stem (0.60-0.63). Contrarily, lower heritability values were estimated for the total number of pods and seeds produced (0.20-0.35), and almost null for the number of branching orders (0.06). Overall, variance component estimates showed a preponderant effect of the environment (Table1). For this reason, the genetic analysis of these traits should be conducted on the singular specific environments separately.

TABLE 1| Estimates of heritability (H^2) of breeding traits in single locations (from Gulisano *et al.* 2022) and across trials. For the analysis of variance across trials, the percentage of variance explained by the different components of variance is reported.

	EC	**PT**	**NL-Sc**	**NL-Wi**	**Across trials**					
	H^2	H^2	H^2	H^2	G	E	Block	GxE	ε	H^2
Plant height	0.72	0.86	0.56	0.57	16.4	45.1	1.9	3.8	32.9	0.82
Flowering time	-	0.93	~~0.24~~	0.87	10.8	86.6	0.2	1.4	0.9	0.88
Branching orders	0.33	0.44	0.46	0.07	0.4	44.5	2.7	0.4	42.9	0.06
Vegetative yield	-	0.54	0.17	0.21	3.5	27.1	5.0	1.5	62.9	0.39
Pods on main stem	0.48	0.42	0.83	0.46	6.2	57.0	0.8	6.6	29.5	0.60
Seeds on main stem	0.46	0.57	0.83	0.53	4.7	70.2	0.3	4.7	20.4	0.63
Pods Total	0.23	0.41	0.69	0.58	1.4	50.0	1.1	9.6	38.0	0.20
Seeds Total	0.21	0.36	0.70	0.64	2.3	58.2	0.9	6.1	32.5	0.35
100 Seed weight	0.67	0.17	0.63	0.24	2.7	39.5	0.2	3.5	54.1	0.33

Total Variance was decomposed in: genotype (G), environment (E), block effects within trial (Block), interaction of genotype by environment (GxE) and Residuals (ε) effects, reported as percentage of the total variance. Traits that were not measured are indicated with a dash. Crossed out traits were discarded for this study.

Genome Wide Association Studies

The GWAS analysis was performed using the R package StatgenGWAS. The Kinship relationship matrix among samples, calculated with the Van Raden method, was included together with the country of origin of the accessions as covariate, to adjust for population

structure. The quantile distribution of the observed p-values versus the expected p-values (QQ plot) confirmed adequacy of the population structure correction.

Due to the large effect of genotype by environment interaction on the phenotypic variation, GWAS was performed analyzing each trial separately, as well as across all the locations. By applying the Bonferroni threshold ($-\log_{10}(p) > 5.52$), a total of 3 significant SNPs were identified to be associated with flowering time, 1 SNP was associated with plant height, 1 SNP with vegetative yield and 3 SNPs with the number of pods on the main stem (Pods MS) (reported in Fig. 2 and Table 2). Contrarily, no significant SNPs were found in association with the remaining phenotypic traits. Single location GWAS for flowering time revealed the association of the SNP M11043, located on chromosome 13, with flowering time in the Portuguese environment. M11034 explained ~6 % of the phenotypic variance observed. In NL-Wi, SNP M7339 on chromosome 8 was found to explain almost 14 % of variation in time to flowering. When the two trials (PT and NL-Wi) were analyzed together, M7339 remained significantly associated to flowering time and explained 11 % of variation across trials. Additionally, M7670 on chromosome 8 was also detected as significant across trial, explaining 16 % of variation in flowering time. Notably, it was possible to observe a similar piling up of SNPs on chromosome 8 also in PT, harboring the same markers detected across trials in association with flowering time (M7670 and M7339) but below the stringent Bonferroni detection threshold (Figure 3). Instead, the SNP M11034 was also associated with the number of Pods MS in the trial located in the Netherlands in 2019 (NL-Sc). For Pods MS, two more SNPs were detected on chromosome 18 (M14832) and chromosome 17 (M14675) in Ecuador and NL-Sc, respectively. About 10% of the plant height variation in Ecuador was ascribable to the SNPs M396 on chromosome 1, while no other SNPs associated with this trait passed the Bonferroni threshold in the other locations, despite the high heritability and genetic variation recorded. However, the visual inspection of Manhattan plots revealed the presence of different regions likely associated with plant height in the other trials (Figure 4). The importance of these SNPs needs to be further investigated considering that Bonferroni is a quite conservative threshold. As an example, an interesting piling up of SNPs right below the threshold was observed on chromosome 16 for PT, while another piling up in association with this trait was observed on chromosome 12 for both NL-Sc and NL-Wi. Finally, 1 SNP associated to vegetative yield was detected on chromosome 12 (M10925), but only in NL-Sc. The SNPs detected as significant are reported in Table 2.

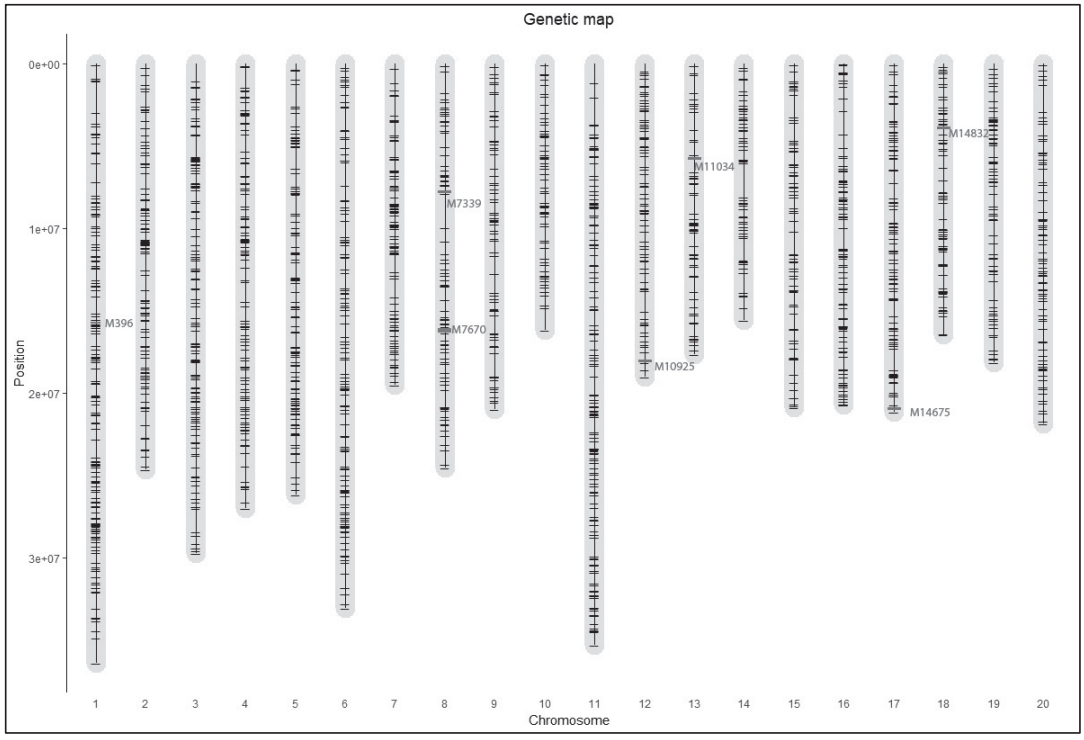


FIGURE 3| Markers distribution on the 20 chromosomes of *L. angustifolius* genome. In red, significant markers detected in association with agronomic traits using Bonferroni threshold (LOD 5.52, experiment wise error $\alpha = 0.05$).

TABLE 2| SNPs found in significant association with plant-architecture and yield related traits for 223 accessions of *L. mutabilis* grown in four different environments.

Traits	Markers	Chr	Position	Environment	LOD	% Var	Effect	N. of genes
Flowering time (days)	M11034	13	5,615,305	PT	6.39	6.35	15.49	14
	M7339	08	7,668,768	NL-Wi	6.69	13.9	9.44	20
	M7339	08	7,668,768	All Trials	5.67	10.6	6.86	20
	M7670	08	16,132,804	All Trials	5.97	16.1	-21.86	1
Plant height (cm)	M396	01	15,690,000	EC	5.99	10.3	14.24	2
Vegetative yield (g)	M10925	12	18,150,269	NL-Sc	5.95	19.3	-168.78	29
Pods on the main stem (number)	M11034	13	5,615,305	NL-Sc	5.99	5.93	-5.67	14
	M14832	18	3,860,231	EC	5.62	10.7	9.02	5
	M14675	17	20,841,361	NL-Wi	6.03	7.00	4.03	16

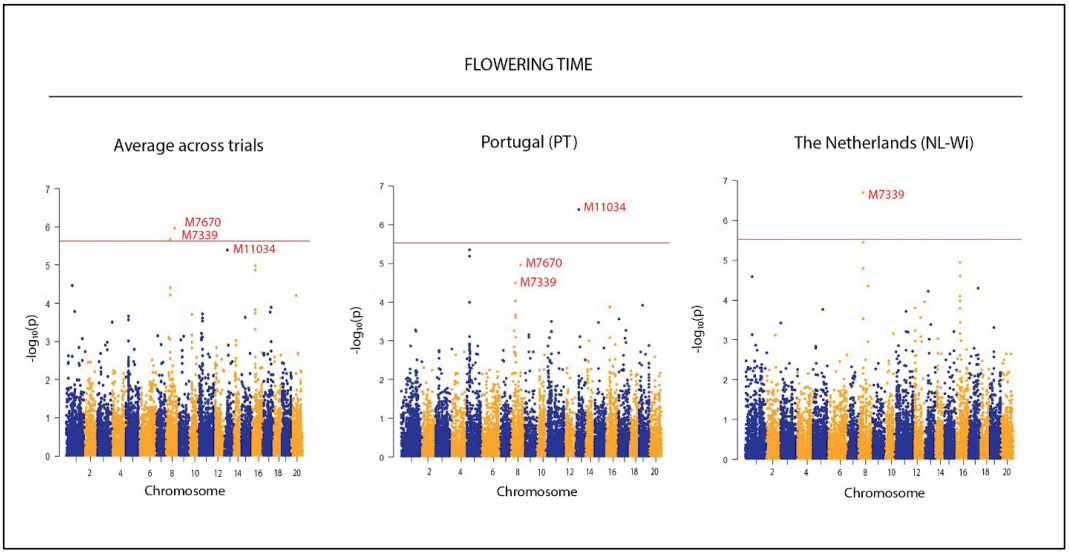


FIGURE 4 Manhattan plots displaying SNP markers-trait associations identified for Flowering time in GWAS using 223 accessions of *L. mutabilis*. The red line indicates the Bonferroni threshold (LOD = 5.52). Common SNPs showed significant association with variation in flowering time across locations, even if not all of them were detected as significant. SNPs that are detected in more than one environment, or in association with more than one trait, are reported in bold. For each marker we report: the position on *L. angustifolius* chromosome (Chr), the trial where the association was detected as significant (Environment), the LOD value of association, the phenotypic variance explained (% Var), the allelic effect on the phenotypic mean of the trait (Effect) and the number of genes found in a window of ± 80 kbp from the marker.

Putative candidate genes

We searched for putative candidate genes in regions close to the peak of the eight QTL detected. An interval of ± 80 kbp around the peak of these QTL was taken as search window for candidate genes as this is about the size of the linkage disequilibrium blocks, hence genes outside this window are not likely to be causative for the QTL effect. All search regions together harbored a total of 86 unique genes on the *L. angustifolius* genome. Moreover, the investigation of other interesting regions previously mentioned as likely associated with plant height, which did not pass the stringent Bonferroni correction, led to the identification of 37 additional genes, mostly related to plant development. The investigation of these genes, based on functional annotation of their orthologs on *Arabidopsis* genome and relevant literature, resulted in a subset of 10 genes as possible genes involved in the control of traits of interest. These genes, listed in Table 3, are known to play a role in regulation of flowering and are involved in different plant growth and development processes. For instance, the identified genes encoding the protein FANTASTIC FOUR 3 (*FAF3*), and the transcription factors LNK3 and GAMYB, are known to contribute to the genetic control of flowering time in Arabidopsis and other species (Wahl et al., 2010). *FAF3*

is located on chromosome 8 at a distance of 77 kbp from the SNP M7399 associated with flowering, while LNK3 and GAMYB are located on chromosome 13 (respectively, 62 kbp upstream and 44 kbp downstream of SNP M11034). A gene encoding an E3-ubiquitin protein ligase RING1 was detected on chromosome 17 in linkage with the SNPs M14675, that showed significant association with Pods MS. Members of the E3 ubiquitin ligases are well known to play a role in plant growth and development. Additionally, important genes involved in plant growth were identified on chromosome 12 and 16. Genetic polymorphism in these genes may be associated with plant height and vegetative yield, and captured by blocks of closely located markers: M12928 to M12945 on chromosome 16 and M10488, M10515 and M10527 on chromosome 12 (Figure 4). These included transcription factors involved in vascular development (*RF2b*), shade-induced plant growth (*HAT3*), transcriptional regulators of flowering time (*Dr1, BIM1*) and genes involved in the gibberellin signaling pathway (*GA3OX1*).

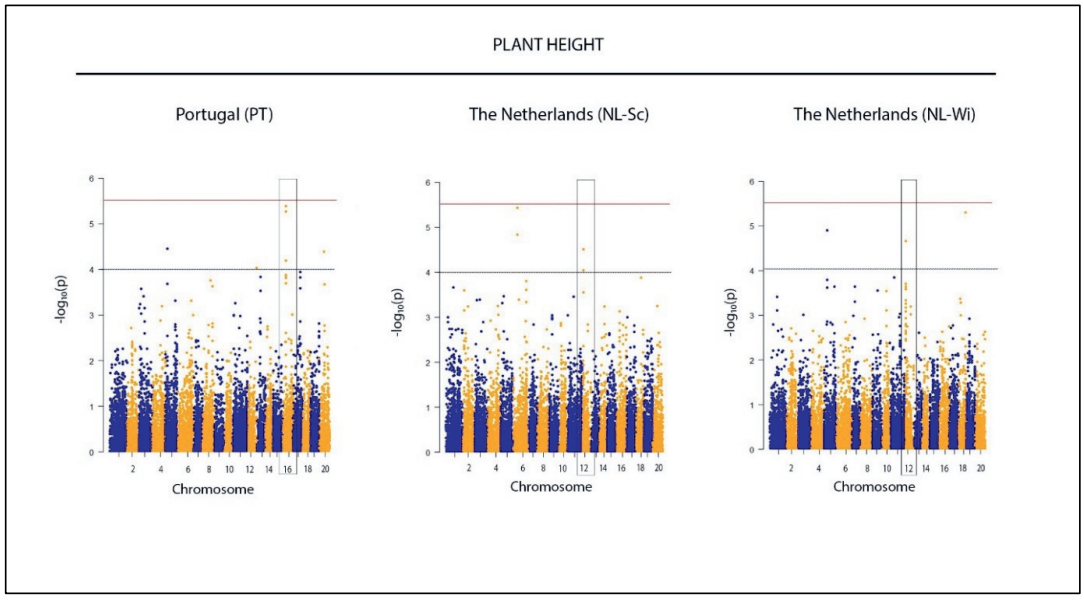


FIGURE 5| Manhattan plots suggestive of a SNP markers-trait association for plant height on chromosome 16 and 12, respectively for *L. mutabilis* growing in Mediterranean (Portugal) and North-Central European cropping conditions (NL-Sc and NL-Wi). The red line indicates the Bonferroni threshold (LOD = 5.52). The black dotted line indicates a suggestive threshold of LOD= 4. The rectangles highlight interesting piling up of SNPs below the threshold of significance, harboring interesting candidate genes related to plant height.

TABLE 3| Functional annotation of the potential candidate genes for plant architecture and yield-related traits in *L. mutabilis*.

Trait	Chr	Candidate genes	Function annotation/ common name
Plant height (PT)	16*	LOC109329023	transcription factor RF2b-like
Plant height (NL)	12*	LOC109362850	homeobox-leucine zipper protein HAT3
Vegetative yield (NL-Sc)	12	LOC109361654	gibberellin 3-beta-dioxygenase 1/ GA3OX1
	12	LOC109362161	histone deacetylase 15 / HDA15
	12	LOC109362682	transcription factor BIM1
	12	LOC109362691	protein Dr1 homolog
Flowering	08	LOC109354086	protein FANTASTIC FOUR 3
Flowering (PT)/	13	LOC109325309	protein LNK3
Pods main stem (NL-Sc)	13	LOC109363186	transcription factor GAMYB
Pods main stem (NL-Wi)	17	LOC109330763	E3 ubiquitin-protein ligase RING1

*interesting genes found in linkage disequilibrium with SNPs detected below the Bonferroni threshold (LOD ≥ 4)

Discussion

Large phenotypic variation for plant height, flowering time, biomass and yield related traits was observed among the 223 *L. mutabilis* accessions studied, indicating that the used panel is suitable for a GWAS study. The large phenotypic variation observed in our population is beneficial to select breeding candidates in order to initiate a genetic improvement program aimed at developing superior varieties adapted to European environments. One of the key factor to promote the cultivation of high yielding *L. mutabilis* in Europe is breeding towards earliness and semi-determinate growth habit (Caligari *et al.*, 2000; Gulisano *et al.*, 2022). In our extensive phenotypic evaluation of this collection, we highlighted the presence of strong correlations between plant height, flowering time and seed yield (Gulisano *et al.*, 2022). These findings are in line with previous observations carried out in other legumes (Annicchiarico *et al.*, 2021, González *et al.*, 2016). Breeding for plant architecture should be tailored to specific growing conditions, considering that environmental conditions highly impact the morphology and yield of *L. mutabilis* (Gulisano et al., 2022). Generally, restricted plant architecture, caused by a limited number of branching orders, is associated with higher seed yield in *L. mutabilis*. In fact, indeterminate growth and extensive branching have a detrimental effect on seed yield, since Andean lupin tends to produce the largest share of seeds on the main stem and on the first branching order (A.A. Sinjushin, 2015; Gulisano *et al.*, 2022). However, from our study emerges that total seed production positively correlates with late flowering and higher vegetative biomass when Andean lupin is cultivated in Mediterranean cropping conditions (Portugal). This observation supports the hypothesis that plants with indeterminate growth, thus more branching orders, can outperform determinate varieties in adverse climates. One possible explanation is that, under high temperature or drought, accessions with indeterminate growth habit may compensate better

for yield loss than determinate types, where pods set is stronger negatively affected by the environmental factors (Martins *et al.*, 2016; Gulisano *et al.*, 2022). However, it is important to report that extensive branching in Mediterranean conditions is already constrained by environmental factors such as water scarcity, hence even plants characterized by an indeterminate growth habit will not produce more than two branching orders, maintaining overall a semi-determinate growth. Yet, the effect that a larger contribution of higher branching orders to seed yield can have on seed quality and composition remains to be further investigated.

Given the importance of breeding for earliness and growth habit in this crop, the collection under study showed enough phenotypic and genetic diversity in flowering time (about one month difference between early and late flowering genotypes), and vegetative yield, to potentially breed for novel early and high yielding lines. The highlighted variation is therefore suitable to support breeding aimed at the introduction of *L. mutabilis* as a protein and oil crop in Europe. For genetic improvement purposes, understanding the genetic architecture of earliness, plants architecture and seed-yield related traits is pivotal. By evaluating strategic breeding parameters including heritability and genotype-by-environment interactions (GxE), we demonstrated that some phenotypic traits of interest are highly influenced by their genetic components. We report high heritability values for flowering (H^2=0.88), plant height (H^2=0.82) and production of pods and seeds on the main stem (H^2=0.6) across the different trials, thus across environments and growing conditions (Table1). Our results are in line with the heritability values reported for plant height, productivity on the main stem and flowering time by previous authors (A. Hardy *et al.*, 1998; Guilengue, Alves*, et al.*, 2020). These findings encourage breeding towards early varieties with restricted plant architecture and improved yield in Europe, since high heritability is always advantageous to increase the expected genetic gain in crop selection.

To the best of our knowledge, this is the first genome-wide association study (GWAS) investigating the genetic architecture of agronomic traits in a diverse collection of *L. mutabilis* accessions. Since genotype by environment interaction can alter the QTL effects, an initial GWAS analysis carried out on single trial data, followed by a genome-wide scan across all locations, was considered the best approach to analyze our data. We have already shown in previous studies that in the absence of a reference genome for *L. mutabilis*, the use of *L. angustifolius* pseudochromosomes assembly can be suitable for genome-wide associations in this species (Chapter 4). In addition, our GWAS approach also shows that, in the absence of pure highly stable homozygous lines, a mapping experiment can be conducted using allele frequencies as marker score, estimated at each marker locus in pooled DNA samples. We estimated allele frequencies by counting the number of reads carrying the reference alleles on the total number of generated reads at a given variant position. In order to ensure a more accurate estimation of allele frequencies, as well as accounting for the presence of alleles of different individuals in the same DNA sample, we have discarded SNPs with a reads depth <45x from the final markers set. This approach of treating markers as a variable with a continuous distribution was already adopted by (Petit *et al.*, 2020), and is a good option for mapping studies in crops where a certain

degree of cross-pollination is expected within the accessions. The Efficient Mixed Model Association (EMMA) approach showed robust performance on the traits and in the population studied. In GWAS, false discoveries are a major concern and spurious associations between phenotypes and testing markers can arise as a consequence of population structure and differences in relatedness among individuals in the tested population (Sul *et al.*, 2018). The association model we used, incorporated kinship and the origin of accessions as covariates to control for spurious association and false discoveries. The QQ-plots generated after the genome scan confirmed the high quality of the GWAS models. The difference between observed and expected −log10(p) values, in fact, showed no inflation and values of inflation factors close to the ideal value of 1 for all the traits.

GWAS analysis identified a total of 3 significative SNPs associated with flowering time (M11034, chromosome 13, position 5,615,305; M7339, chromosome 8, position 7,668,768; M7670, chromosome 8, position 16,132,804), 1 SNP associated with plant height (M369, chromosome 1, position 15,690,000), 1 SNP associated with vegetative yield (M10925, chromosome 12, position 18,150,269), and 3 SNPs associated with number of pods on the main stem (M1134, chromosome 13, position 5,615,305; M14832 on chromosome 18, position 3,860,231; M14675, chromosome 17, position 20,841,361). In agreement with the idea that the genetic control of agronomic traits might differ in response to different environmental conditions, our study points out that the association of genomic regions with phenotypic traits can be location specific, thus different QTL can be detected in different environments. Indeed, most of the SNPs reported were identified in single trials only, while the markers M7339 and M7670 were the only ones detected across trials as associated with flowering time. Taking into account that the Bonferroni correction method is a very conservative methodology to set a threshold *P*-value (Kaler *et al.*, 2019), we have also investigated some genomic regions where SNPs were piling up in the QQ plots, even if they were not passing the stringent Bonferroni threshold. We therefore report on likely association on chromosome 12 and chromosome 16 for plant height in the Netherlands (NL-Sc and NL-Wi) and in Portugal (PT), respectively.

Flowering time is a quantitative traits and its genetic control has been investigated in several crops, including *L. albus,* where it has been described as quantitative and under the control of major QTL and numerous regulatory genes, including *FLOWERING LOCUS T (Fta1), CONSTANS (CO), FY, MOTHER OF FT AND TERMINAL FLOWERING LOCUS 1 (TFL1),* but also of genes related to the response to photoperiod and vernalization, as well as regions involved in hormone signaling pathways (i.e, gibberellin) (Rychel *et al.*, 2019). In our panel, we did not detect these major candidate genes conserved in several crops. However, the variation in flowering time in our panel resulted associated with two large QTL on chromosome 8 (M7339 and M7670) across locations, and associated with a QTL on chromosome 13 (M11043) in PT. Given that M7339 and M7670 respectively accounted for ~11 % and 16 % of the phenotypic variation observed across trials, it is possible to hypothesize the presence of major QTL in their surrounding genomic regions. Our investigation highlighted the presence of the

gene FANTASTIC FOUR 3 (*FAF3*) at a distance of 46 kb apart from M7339, thus located in a window where the putative casual mutation is expected to be in linkage disequilibrium with this proximal marker. *FAF3* is a member of the FANTASTIC FOUR (FAF) genes family, which contains four members in *Arabidopsis* (*FAF1-FAF4*), dynamically expressed throughout development (Wahl *et al.*, 2010), and mainly involved in the modulation of the shoot meristem size. Wahl et al. 2010 reported FAF genes as redundant in function, indicating that all the four proteins may perform a very similar role. Interestingly, the FAF genes, are expressed in flowering buds and inflorescence, with an increased expression during the transition to flowering. Moreover, a regulatory interaction has been reported between *FAF2* and *FAF4* with the WUSCHEL (WU)-CLAVATA (CLV3) signaling pathway that plays a central role in regulating stem cell proliferation and differentiation in crop plants, supporting organogenesis in the floral meristem through a complex molecular pathway (Wahl *et al.*, 2010). Notably, studies on *Arabidopsis thaliana* report *FAF1/FAF2* and *FAF3/FAF4* to be recently duplicated paralogs (Blanc *et al.*, 2004) that could therefore be characterized by similar structure and functions. Unfortunately it was not possible to detect any annotated and characterized gene in the genomic region surrounding M7670, hence further studies are needed to elucidate which gene is responsible for the variation in flowering time associated with this SNP. In addition to M7339, we report the finding of SNP M11043 on chromosome 13 as associated to flowering time in Mediterranean growing conditions. M11043 is located in a genomic area that harbors two genes potentially involved in the regulation of flowering. The gene *GAMYB* is located at a distance of 44.2 kbp from the detected marker, while the gene NIGHTLIGHT–INDUCIBLE AND CLOCK-REGULATED 3 *(LNK3)* at a distance of 62.3 kbp. *GAMYB* is described as a regulator of flower induction, via the transcriptional activation of the LEAFY (*LFY*) gene (Gocal *et al.*, 1999; Y. Zhang *et al.*, 2020), that results in early flowering both in dicots and monocot (He *et al.*, 2000; Weigel *et al.*, 1995). Instead, members of the family of NIGHTLIGHT–INDUCIBLE AND CLOCK-REGULATED (*LNK1* and *LNK2*) genes are known to control photomorphogenic and photoperiodic responses, as well as circadian rhythms (Rugnone *et al.*, 2013). Recently, Li et al. 2021, obtained an early flowering line in soybean by targeted mutagenesis of the four *LNK2* genes, that molecularly interact with major flowering genes, showing that target breeding on these genes can contribute to soybean breeding in high-latitude European environments (Zhaobo Li *et al.*, 2021). Despite we report on *LNK3*, that was not previously described as involved in flowering, it will be beneficial to further investigate its role in *L. mutabilis*, based on the strong significant association detected, and the fact that the polymorphism detected explains ~6% of the flowering variation recorded.

The genetic analysis carried out in this study showed that plant height and vegetative yield are likely affected by QTL-by-environment interaction, since the detected SNPs and the respective candidate QTL in linkage with them appeared to be environment-specific. Hence, different transcriptional regulators can similarly affect the same traits in response to different environmental stimuli. In particular, plant height showed association with the SNP M396 (chromosome 1, position 15,690,000) in Ecuador, but a piling up of SNP (not significant based

on Bonferroni) was instead observed on chromosome 16 in Portugal, and on chromosome 12 in the Netherlands. On chromosome 1 we did not report any relevant candidate genes linked to M396 and plant height. Contrarily, we report the finding of a transcription factor *RF2b-like* and a homeobox-leucine zipper protein *HAT3*, both harbored in the genomic regions where we visualize an interesting piling up of SNPs, that likely resulted as false negative when applying Bonferroni threshold to detect significant associations. These genes are respectively involved in vascular development and shoot tissue organization (*RF2b)* and involved in shade-induced plant growth (*HAT3)*, and can both lead, through different mechanisms, to the development of semi-dwarf varieties (Dai, Zhang, Chen, & Beachy, 2004; Bou-Torrent et al., 2012).

The association between genetic variants and vegetative yield was only significant in North-Central European summer conditions, were the SNP M10925 (chromosome 12, position 18,150,269) was detected. M10925 showed a strong association and it is able to explain a large share (~20%) of the phenotypic variation. This suggest the presence of a major QTL, considering that vegetative yield is a quantitative and complex trait. Harbored in these area, we identified 4 different candidate genes: *giberellin 3-beta-dioxygenase1 (GA30X1)*, *histone deacetylase15 (HDA15)*, transcription factor *BIM1* and *D*r1 homolog, which are gene involved in the gibberellin pathway (*GA30X1*) and regulation of flowering time (*BIM, Dr1*). On one hand, GA30X1 is involved in the gibberellin pathways as the enzyme catalyzing the final step of the synthesis of bioactive gibberellin during vegetative growth, and its negative regulation might directly lead to substantial decreases in biomass production and to the development of semidwarf types (Hu et al., 2008). On the other hand, transcriptional regulators of flowering time can promote flowering in response to hormonal shifts or environmental stimuli and, as a consequence of the strong interplay between flowering time and growth habit in legumes, indirectly act on biomass production and vegetative yield. For example, the transcription factor *BIM1*, part of the bHLH family of transcription factors, can lead to the inhibition of floral transition in Arabidopsis by transduction of hormonal brassinosteroid signals to the activation of FLOWERING LOCUS C (*FLC*) and consequent floral repression, promoting in turn vegetative growth. Conversely, defects in *BIM1* can lead to early floral transition (Zicong Li *et al.*, 2018). Instead, *Dr1* and *HDA15* play important roles in determining plant response to different environmental stimuli, such as drought and elevated ambient temperature (Zotova *et al.*, 2019; Shen *et al.*, 2019). For example, it is suggested that Dr1 protein can bind and deactivate genes sensing environmental cues for drought, releasing in turn the activity of vernalization and flowering regulators (*Vrn1* ad *FT1*). This mechanism has been described as a successful drought escape strategy in wheat (Zotova *et al.*, 2019), and could be linked in *L. mutabilis* to other successful escape strategies like accelerated seed maturation and pod filling which rely on a preferential partitioning of nutrients towards reproductive growth, at the expense of vegetative growth.

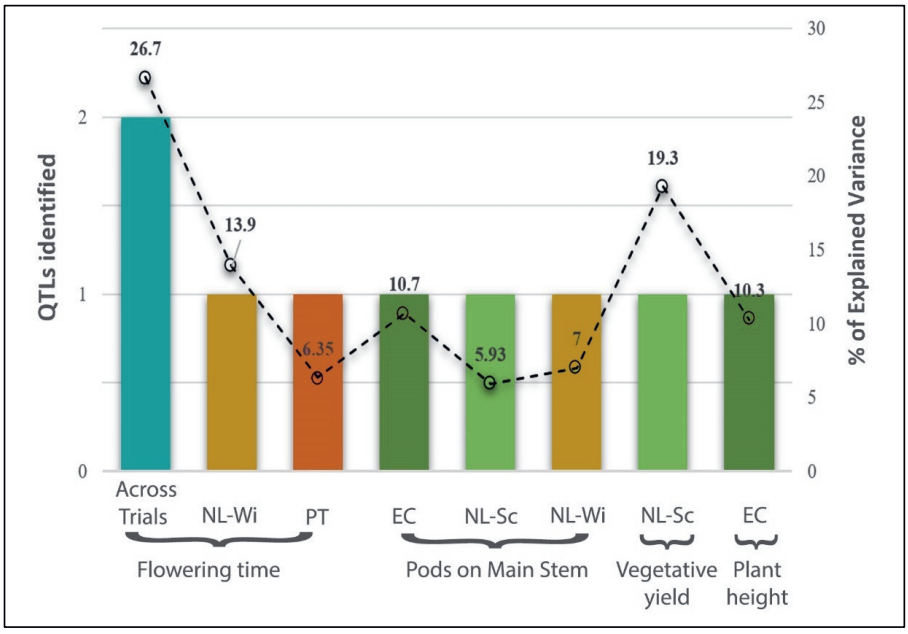
Finally, single QTL location-specific were also detected in association with number of pods produced on the main stem (Pods MS). M11034 (chromosome 13, position 5,615,305) was detected in the NL-Sc, M1483 (chromosome 18, position 3,860,231) in EC and M14675

(chromosome 17, position 20,841,361) in NL-Wi. On chromosome 17, we highlighted the presence of the E3 ubiquitin ligases *RING1* gene in linkage disequilibrium with the respective significant marker. E3 ubiquitin ligases have been documented to play an important role in the regulation of plant growth and development, such as seed dormancy and germination, root growth and flowering time control, as well as regulation of several abiotic stress responses (H. Zhang *et al.*, 2015; Shu *et al.*, 2017). Increasing numbers of studies have documented the key role of the different types of E3s, including E3-RING being involved in seed biology, root elongation, flowering time control, light response, ABA signaling transduction and response to drought and salinity stresses.

Conclusions

In this study, we explored the phenotypic and genetic diversity of a wide collection of *L. mutabilis* accessions to investigate for the first time the genetic architecture of agronomic traits in this species, using a GWAS approach. The association of genomic regions with phenotypic traits was found to be environment specific for some traits for which location specific QTL were found (e.g. for vegetative yield or number of pods on the main stem). In addition, our study allowed an analysis of the stability of the effect of QTL under different environments. Average values across environments, indicated two important QTL for flowering time on chromosome 8 explaining 11 and 16 % of the variation observed in flowering time. We also found QTL associated with flowering time or vegetative yield in only one or two of the environments but not in all. For example, we report a QTL on chromosome 13 associated with flowering time in Mediterranean environments and a QTL on chromosome 12 associated with vegetative yield in North-European conditions. After analyzing the annotated genes in the genome of lupin for the genomic regions spanned by these QTL, we found several genes and transcription factors (*FAF, GAMYB* and *LNK*) with known functions relating to the regulation of major pathways influencing flowering and determinate/indeterminate growth in other species. In other genomic regions around QTL we found genes acting on these pathways in response to hormonal and environmental cues (*GA30X1*, *BIM1*, *Dr1*, *HDA15*, *HAT3*). Alleles for the QTL can be detected on the basis of the DNA polymorphism and thus directly be used to initiate marker-assisted selection which can speed up the development of *L. mutabilis* varieties adapted to specific cropping conditions. If the underlying genes responsible for the QTL can be identified, a wider search for functional alleles becomes possible. Additionally, techniques like mutation breeding or gene editing could be directed at such genes. This study can serve as a starting point for such deeper investigations of the mechanism regulating the intricate system interconnecting plant development and phenology in this species.

Supplementary Figures and Tables



SUPPL. FIG 1|. Survey of the QTL identified for agronomic traits in a collection of 223 *L. mutabilis* accessions. Environment-specific QTL were identified in a single location (PT, EC, NL-Sc or NL-Wi), while stable QTL across environments were identified using average values across environments (Across Trials). Coloured bars indicate total number of QTL identified per trait in our analyses. Circular points refer to the cumulative proportion of variance explained by all identified SNPs for a given trait.

6

Phenotypic diversity of protein and oil content, and oil composition in *L. mutabilis* seeds

Agata Gulisano[1], Wilrik Aanroi, Annemarie Dechesne[1], Maria-João Paulo[2], Luisa Trindade[1]

[1] Wageningen University & Research Plant Breeding, Wageningen University, Wageningen, Netherlands,
[2] Wageningen University and Research Biometris, Wageningen Research, Wageningen, Netherlands
[3] INIAP, Estación Experimental Santa Catalina, Quito, Ecuador,

To be submitted.

Abstract

Lupinus mutabilis seeds distinguish themselves from several other common legume grains for a high protein and fat content, and very low starch. Preliminary analysis of few genotypes have reported an average content of 44% protein and 18% oil in these seeds, presenting the Andean lupin as an interesting alternative to soybean. The introduction of this crop in Europe requires, however, the development of varieties that are adapted to the European conditions. Fast and effective determinations of oil content, fatty acid composition and protein content are key in this process, both to drive the selection of elite material and to assist and speed up breeding programs. In this study, we assess the use of Near-infrared reflectance spectroscopy for a high-throughput characterization of *L. mutabilis* seeds quality on a panel of 220 accessions growing in Mediterranean conditions. Predictive models were developed for protein content, oil content, and fatty acid composition, and were then evaluated through cross-validations and external validations. Prediction models with reliable estimations of protein content, oil content and fatty acid composition (stearic, oleic, linoleic, linolenic and behenic acids) were obtained and applied to characterize seed quality in the panel, providing valuable insights into the variability of these traits across accessions. Bolivian and Ecuadorian material emerge as characterized by the highest oil (~15% dm) and protein (~44% dm) concentrations in the seeds respectively. Furthermore, due to a fatty acid profile rich in oleic (41 %), linoleic (34 %), palmitic (11 %) and stearic (7 %) acids, and low in linolenic acid (~2.5 %), *L. mutabilis* oil appears suitable both for edible and industrial purposes, and well suited for biodiesel production.

Keywords: Andean lupin, oil content, protein content, fatty acid composition, NIRS

Introduction

The development of sustainable and competitive crops able to meet the European ever-growing demand for protein, oil and biomass is key to the achievements of the European Green Deal targets. The need to become less dependent on the import of soybean and to create a more sustainable agriculture in Europe, has put in the spotlight many pulses and protein-rich oilseeds including lupins. Lupin seeds are a rich source of high-quality proteins, lipids, as well as compounds offering potential health benefits (Arnoldi *et al.*, 2011; Czubinski *et al.*, 2021; Hamama *et al.*, 2004; Muzquiz *et al.*, 2012; Ruiz-López *et al.*, 2019; Wang *et al.*, 2008). Many studies have documented that the nutritional quality of lupin seeds is comparable to soybean. Furthermore, lupin species also holds the advantage of having a high potential in the reclamation of poor and marginal lands for agriculture and of not facing the GMO opposition often associated with the use of soybean (Czubinski *et al.*, 2021; Gulisano *et al.*, 2019). The main lupin species currently cultivated in Europe and worldwide are the white lupin (*L. albus*), yellow lupin (*L. luteus* L.) and narrow-leafed lupin (*L. angustifolius* L.). Total protein content in these species varies in the range of 33.9-42.2 g/100 g dm and compares well with protein content of soybean (43 g/ 100 g dm), while fat content varies between 5.5-11.2 g/100 g dm and is lower than that of soybean, 19.8 g/100 g dm.

The past twenty years, have seen an increasing interest in lupins and a strong rise in European cultivated area and production of lupins (FAOSTAT, 2022). However, the success of these species in replacing soybean has been hampered by a low oil content in the seeds and limited adaptation to a narrow range of agroclimatic conditions (Gulisano *et al.*, 2019). *Lupinus mutabilis*, an emerging lupin species that could help overcoming these limitations, has been introduced only last year to the European market (van Haren and van Dinter, 2021). *Lupinus mutabilis*, is a species native to the Andes and still cultivated only in some parts of South America (Lucas *et al.*, 2015), but the high protein and fat content of its seeds and its adaptation to low input farming in temperate climates make it potentially superior than the current plant-based sources of protein and oil in Europe (Gulisano *et al.*, 2019). Crude protein content has been reported to range between 32-52.6 g/100 g dm within *L. mutabilis* species, while oil content ranges between 13-24.6 g/100 g dm (F.E. Carvajal-Larenas *et al.*, 2016). This wide range of variation in protein and oil content can be associated with both genetic and agronomic factors. Previous authors have reported a wide genetic variability for this species, illustrating adaptation to microhabitats as a result of natural selection (Haq, 1993). This variability has been mainly noted in plant shape, vegetative development, flowering time and seed content of oil, protein and alkaloids (Guilengue *et al.*, 2019; Gulisano *et al.*, 2022; Haq, 1993; Martins *et al.*, 2016).

In the past twenty years, numerous efforts have been made to establish *L. mutabilis* as a competitive crop in South America and, at the same time, to introduce this plant to Europe. In Europe, *L. mutabilis* has become the subject of many research programmes, mainly focusing on the adaptation to North-European and Mediterranean environments and on the development of

early varieties characterized by high seed yield and a determinate architecture (Caligari *et al.*, 2000; Guilengue *et al.*, 2019; R Galek, 2010; R Galek *et al.*, 2007). Measurements of seed protein and oil content are however vital to these projects, to drive the selection of starting material required in breeding programs. Recently, Czubinski and co-workers (Czubinski *et al.*, 2021) reported the first detailed characterization of *L. mutabilis* seeds obtained from an ongoing experimental breeding project in Poland. They report a total protein and fat content of 44.7 % dm and 15.4 % dm respectively, and highlight the presence of healthy compounds as oligosaccharides, tocochromanols, carotenoids, phytosterols and flavonoids suggesting the application of *L. mutabilis* seeds for the production of food with health-beneficial effects (Czubinski *et al.*, 2021). However, no research has aimed so far to characterize the genetic variability associated with protein and oil content of South American accessions cultivated in European environments. Considering the wide genetic variability of this species, the evaluation of native germplasm might offer great potential for improving seed quality in diversified environments.

Thus, the present study aims to characterize protein content and oil content and composition in a large panel of 220 *L. mutabilis* accessions growing in Mediterranean cropping conditions. The panel was assembled to harbor as much genetic variation as possible, including both material gathered from the Andean region (Bolivia, Ecuador, Peru and Chile) and material selected from breeding programs in Europe and Belarus. The use of screening methods (NIRS) and the development of predictive models for seed protein and oil content and oil composition in *L. mutabilis*, allowed an high throughput characterization of the collection providing vital information for the selection of valuable breeding material.

Material and Methods

Plant Material

Protein and oil content of seeds was evaluated in a wide panel of 220 genetically diverse *L. mutabilis* accessions. The panel included mostly landraces, varieties and wild material from the centre of origin of the species, the Andean region, provided by the Instituto Nacional de Investigaciones Agropecuarias of Quito (INIAP, Ecuador). Additionally, breeding material was provided by European breeding institutes, namely the Instituto Superior de Agronomia (ISA, Lisbon, Portugal) and the Julius Kühn-Institut (JKI, Quedlinburg, Germany). In detail, the panel included accessions from Ecuador (94 accessions), Peru (65), Bolivia (15), Chile (1), and of unknown origin (17) and lines resulting from breeding programs in Belarus (6), Portugal (18) and Germany (4). The panel was assessed in Portugal, between November and May 2019, as previously described in Chapter 3. The field trial had a randomized complete block design in three replicates, where each experimental unit included 20 plants of the same accessions. Seeds were harvested fresh from the 6 central plants of each block, dried under room conditions and then seed yield (g/plant) and the weight of 100 seeds (g) were scored.

Near infrared reflectance spectroscopy (NIRS)

Each sample was composed of 60 seeds, comprising 10 seeds randomly selected from each of the 6 central plants of each block. Three biological replicates per accessions were included in the experiment, corresponding to the three blocks in the field trial. Each sample was milled using an IKA® A11 Basic Analytical mill, and the resulting flour immediately scanned using a Foss DS2500 near-infrared spectrometer (FOSS, Co. LLC, Denmark). Because of missing plots, 641 out of 675 samples were scanned using NIRS. The remaining 34 samples were either not present or too low in volume to be scanned. Samples were scanned using a small ring cup filled with ground sample and absorbance (Log 1/R) was recorded in the spectral range of 400 to 2500 nm, at 2 nm intervals. Each spectrum was averaged over 7 consecutive scans using ISI-Scan software (FOSS). The WinISI 4.9 statistical software (FOSS) was used for spectral and reference data acquisition, pre-treatment and calibration. Raw spectral data were transformed by using Standard Normal Variate (SNV) and de-trending procedure to eliminate scattering distortions (Barnes *et al.*, 1989). The absorbance data (Log 1/R) was calibrated against wavelength using the second derivatives of transformed spectra, in order to overcome distortions (Infrasoft International, 1999; Vaknin et al., 2011) (see Figure 1). A subset of samples was selected for calibration based on the variation of the NIR spectra and biochemically analysed to develop prediction models for seed protein and oil concentration and oil composition. Respectively, a subset of 100 samples was selected for protein calibration and a subset of 30 samples for oil calibration. After the removal of outlying values, two subsets of 94 and 28 samples were obtained respectively. Calibrations equations were developed based on the threated data, by application of modified partial least square (MPLS) regression method by WinISI software (Infrasoft International, 1999). The coefficient of determination of the calibration (R_c^2), the standard error of calibration (SEC) and the standard error of cross validation (SECV) were used as indicators to select the best equation for prediction. The predictive capability of the calibration equation was assayed using the leave one out cross-validation algorithm, where each sample is removed from the calibration set and a new calibration model is developed and used to predict the sample not in that model. The ability of the calibration equations to predict oil and protein content in a subset of 25 randomly chosen external samples was assayed by external validation (Suppl. Fig.1). The validation coefficient of determination (R_c^2 /R_v^2), the standard deviation of the reference values (SD), the standard errors of calibration (SEC) and of cross validation (SECV), and the standards errors of prediction of external validation corrected for the bias (SEP(C)) are reported in Table 1 and served as indicators of prediction quality.

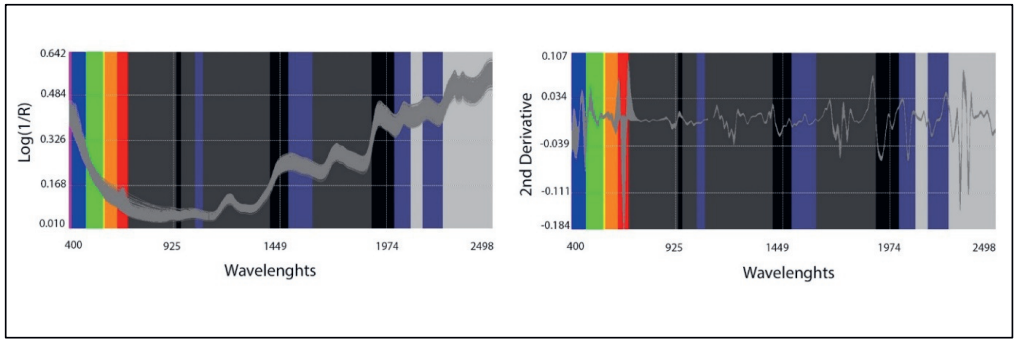


FIGURE 1| NIR spectra of the *L. mutabilis* panel (n=641). On the left VIS-NIRS reflectance spectra and on the right the second derivative pre-processed VIS-NIRS spectra.

Protein content

For the subset of selected samples for calibration, 250 mg of ground seeds were analysed for total N content at the Food Process Engineering Lab (Wageningen University) using the Dumas Method (Elementar, rapid N exceed). Samples were run in triplicate, using 250 mg of aspartic acid as standard. The "Standard Method" where the combustion tube is at 960 °C and the oxygen is dosed for 80 s at 170 mL^{-1} was followed. A factor of 5.37 was used to convert total N to protein, as proposed by Mosse (Mosse, 1990) for *Lupinus albus*.

Oil content

Oil content was determined by solvent extraction using n-hexane. Five hundred mg of ground sample were mixed with 5 ml of hexane and incubated at 40˚C for 30min. After centrifugation, the supernatant containing oil and hexane was separated from the remaining sample and the hexane evaporated using the RapidVap Vacuum Dry Evaporation System (Labconco Co). Oil content was determined by weighing the residues obtained and reported as % of the dry matter content of the initial sample.

Oil composition

The fatty acid fraction was extracted from 75 mg of ground *L. mutabilis* seeds and the compounds of fatty acid methyl esters (FAMEs) were analysed and identified using gas chromatograph–mass spectrometry (GC–MS; column/DB-23; Agilent Technology, Santa Clara, CA). The exact weight was recorded and then 1 ml of hexane and 60 µl of 4.25M KOH/MeOH were added to the samples, followed by vortexing and heating at 60 ˚C for 10 minutes. After cooling to ambient temperature, FAMEs were recovered as supernatant by centrifuging at 3,000 rpm for 5 minutes. The supernatant was then injected (1 µl) into the GC-MS system with a split ratio of 1:50; the oven condition was 190 °C for 10 min, with a ramp to 220 °C for 5 min with 5 °C increments

per minute. The carrier gas was ultrahigh-purity helium, with a flow of 1ml/min. The FAMEs were identified by comparing their retention time with those of standards, and peak areas were quantified using the Chemstation program (Agilent Technologies UK LTd., Stockport). The concentrations of individual fatty acids were calculated after normalizing the peak areas for the initial weight of the sample (dry matter weight) as proportion of the total area of FAMEs.

Statistical analyses

Statistical analysis were performed in R software. The effects of diverse genotype and diverse origin of the genotype on predicted seed protein content, oil content and oil composition were subjected to analysis of variance (ANOVA) and post hoc Tukey Kramer tests. Pearson correlation analysis was used to evaluate the relationship between the total protein and oil concentration, and the oil composition in the seeds of all accessions pooled together.

Results

Predicted seed protein and oil concentrations significantly varied between *L. mutabilis* genotypes ranging from 37.1 to 45.2 % and from 10.5 to 18 % respectively. The content of oil in seeds was negatively correlated to the content of protein (r= -0.83). Furthermore, oil content was significantly lower in genotypes from Ecuador and unknown origin (~12.3 %), while the highest content was measured in accessions from Bolivia (15.2 %), followed by breeding material from JKI and genotypes from Peru (~14.7 %). Conversely, protein content was significantly higher in accessions from unknown origin and Ecuador (43-44%), and lower in seeds from other sources.

Figure 1 shows the VIS-NIRS spectrum of all scanned 641 *L. mutabilis* seed samples and their distribution after pre-processing of the data using the second derivative. The seeds' total reflectance spectrum appears to be highly varied along the entire wave lengths range of 400 to 2500 nm. The principal component analysis of the calibration samples' spectra (PCA, Figure 2), shows that the small calibration set seems to cover a range of variation similar to the large sample set, by including samples that appeared uniformly distributed across the range of observed values. The use of these two calibration sets for the estimation of protein content delivered in both case highly linear prediction models, characterized respectively by an R_c^2 of 0.96 for the model based on 28 samples and an R_c^2= 0.87 for the model based on 94 calibration samples (Table 1). The external validation was also slightly lower for the model based on a larger calibration set (R_v^2=0.64). Being $R^2 = 1 - \left(\frac{SEC}{SD}\right)^2$, at a lower SEC compared to SD, will correspond an improved R^2. Despite the lower R^2 values for the prediction model developed using the larger calibration set (n=94), a slightly higher accuracy of this model is shown by the smaller difference between SEC and SECV, which should generally not differ more than 20% (Shenk et al., 1992) (see Table 1). However, since the predicted values of seed-protein concentrations based on the two NIRS model were found to be highly linear (Figure 2, r =0.92), it was decided to develop

the prediction models for oil concentration and fatty acid compositions analysing biochemically only a calibration set of 28 samples, hence reducing considerably time and labour. The NIRS-based predictions of seed oil and fatty acid compositions were fairly highly linear related to the measured values, as shown by the values of external validation (Table 2). However, fatty acids present in smaller ratios were characterized by a smaller level of linearity in the prediction models for palmitic, palmitoleic and arachidic acids and it was therefore not possible to predict them (Table 2). Margaric acids was detected only in 17 samples, which were too few to be able to develop a reliable prediction model for this trait.

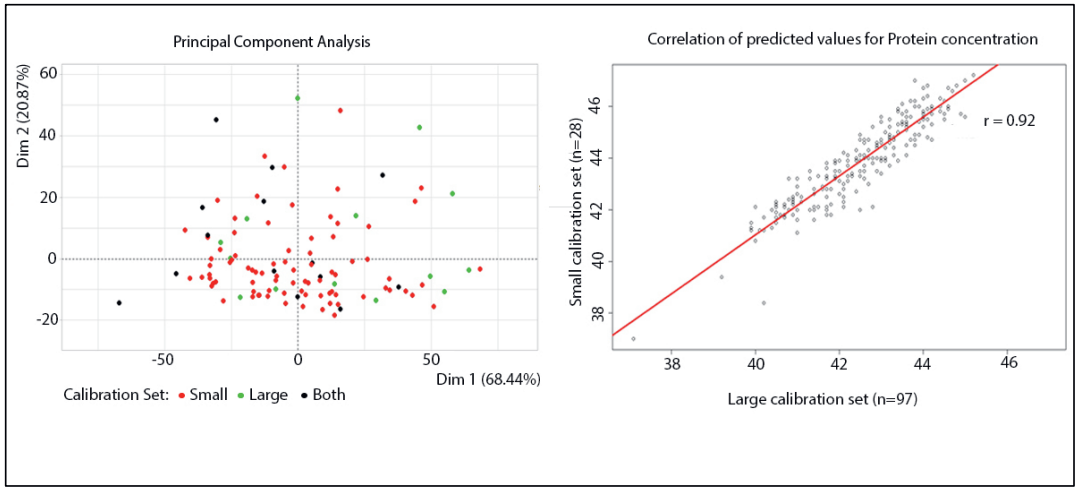


FIGURE 2| On the right, PCA scores plot (PC1 vs PC2) for VIS-NIRS spectra of the two calibration samples sets used in this study for the estimation of protein concentration *in L. mutabilis* seeds. In green, samples belonging to the small calibration sets (n=28); in red, samples belonging to the large calibration set (n=94); in black, samples present in both datasets. On the left, a high correlation (r = 0.92) is shown between the predicted values obtained developing two different predictive models based on the two calibration sets (small vs large) for estimation of seed protein concentration in 220 *L. mutabilis* accessions.

TABLE 1| Cross validation and external validation statistics in the development of calibration equations for Oil concentration (%) and protein concentration (%) in *L. mutabilis* seeds.

	Reference values				Calibration performance			External validation	
	n	Mean	SD	Range	R_c^2	SEC	SECV	SEP(C)	R_v^2
Protein concentration (%)	94	42.70	1.73	37.5-47.9	0.872	0.61	0.71	1.02	0.64
Protein concentration (%)	28	43.35	2.68	38.4-47.1	0.988	0.29	0.70	0.84	0.76

TABLE 2| Cross validation and external validation statistics in the development of calibration equations for oil content (% dm) and individual fatty acids composition (%) in *L. mutabilis* seeds.

	Reference values				Calibration performance			External validation	
	n	**Mean**	**SD**	**Range**	**R_c^2**	**SEC**	**SECV**	**SEP(C)**	**R_v^2**
Oil concentration (%)	28	13.44	2.38	9.0-17.9	0.97	0.42	0.65	0.78	0.81
Palmitic C16:0	28	11.23	0.97	9.53-12.88	0.43	0.57	0.65	0.60	0.34
Palmitoleic C16:1	28	0.23	0.06	0.15-0.37	0.47	0.02	0.02	0.03	0.25
Margaric C17:0	17	0.04	0.04	0-0.10	-	-	-	-	-
Stearic C18:0	28	7.40	1.14	4.74-10.67	0.95	0.17	0.44	0.81	0.66
Oleic C18:1 cis	28	40.84	3.86	35.25-48.9	0.97	1.93	2.17	2.72	0.51
Oleic C18:1 trans	28	0.95	0.09	0.73-1.16	0.76	0.04	0.04	0.15	0.41
Linoleic C18:2	28	34.93	3.39	29.64-42.4	0.98	1.46	1.69	1.95	0.68
Linolenic C18:3	28	2.48	0.24	2.00-2.97	0.82	0.06	0.09	0.11	0.72
Arachidic C20:0	28	0.84	0.13	0.58-1.20	0.67	0.02	0.05	0.12	0.18
Behenic C22:0	28	1.07	0.16	0.81-1.45	0.71	0.05	0.06	0.09	0.40

GC profiling of the oil showed that four fatty acids – oleic, linoleic, palmitic and stearic- were the most abundant in *L. mutabilis* seeds, accounting for 94 % of the oil in seeds from all the tested genotypes. The two unsaturated fatty acids, oleic and linoleic, accounted for the majority of the seed oil composition, 41.4 and 34 % of FAMEs respectively. The two saturated fatty acids, palmitic and stearic, accounted instead only for 11.2 and 7.4 % of the seed oil, respectively. Following, linolenic, behenic, oleic (trans), arachidic and palmitoleic acid accounted only for minor fractions of seed oil, between 0.84 and 2.5% (see Reference values, in Table 4). Whereas contents of most fatty acids were relatively similar for most seed sources and genotypes, those of oleic (cis), linoleic and stearic acid varied more, as shown by the larger ranges of variation characterizing these fatty acids across different origins of the accessions and across all genotypes (Figure 3). Oleic acid (cis) content was significantly higher only in accessions from Peru (44.07 %), where it varied between 38.37 to 50.81 %, showing the highest content in genotype LIB126. Linoleic acid content was instead higher in lines from Belarus (36.58 %), and significantly different from accessions from Peruvian and unknown origin, which had the lowest content (comprised between 32.5-34.6 %). The largest variation in linoleic acid content was observed in material from Bolivia (from 29.2% to 42.7 %), where accession LIB119 was characterized with the highest content of linoleic acid.

Significant correlations were detected between the different fatty acid components and the protein and oil content of seeds, and are displayed in Figure 4. The concentrations of linoleic and

oleic (cis) acids were negatively correlated, with a correlation coefficient of r= -0.85. The oil content was positively correlated to the content of oleic acid (cis) (r= 0.67) and negatively correlated to all the other fatty acids, including linoleic acid (r = -0.8). The content of linoleic acid was also positively correlated to the total protein content (r = 0.6), contrarily to oleic acid which has a strong negative correlation with protein content (r = -0.5). Higher correlations between fatty acids and protein concentrations were reported for oleic acid (trans) and behenic acid, which appear to positively correlate with protein (r = 0.8-0.9) and negatively with oil concentration (r = -0.9). Notably, no correlation was observed between protein and oil concentrations and seed weight; while a slightly positive correlation (r = 0.2) between yield and oil concentration in the seed was detected.

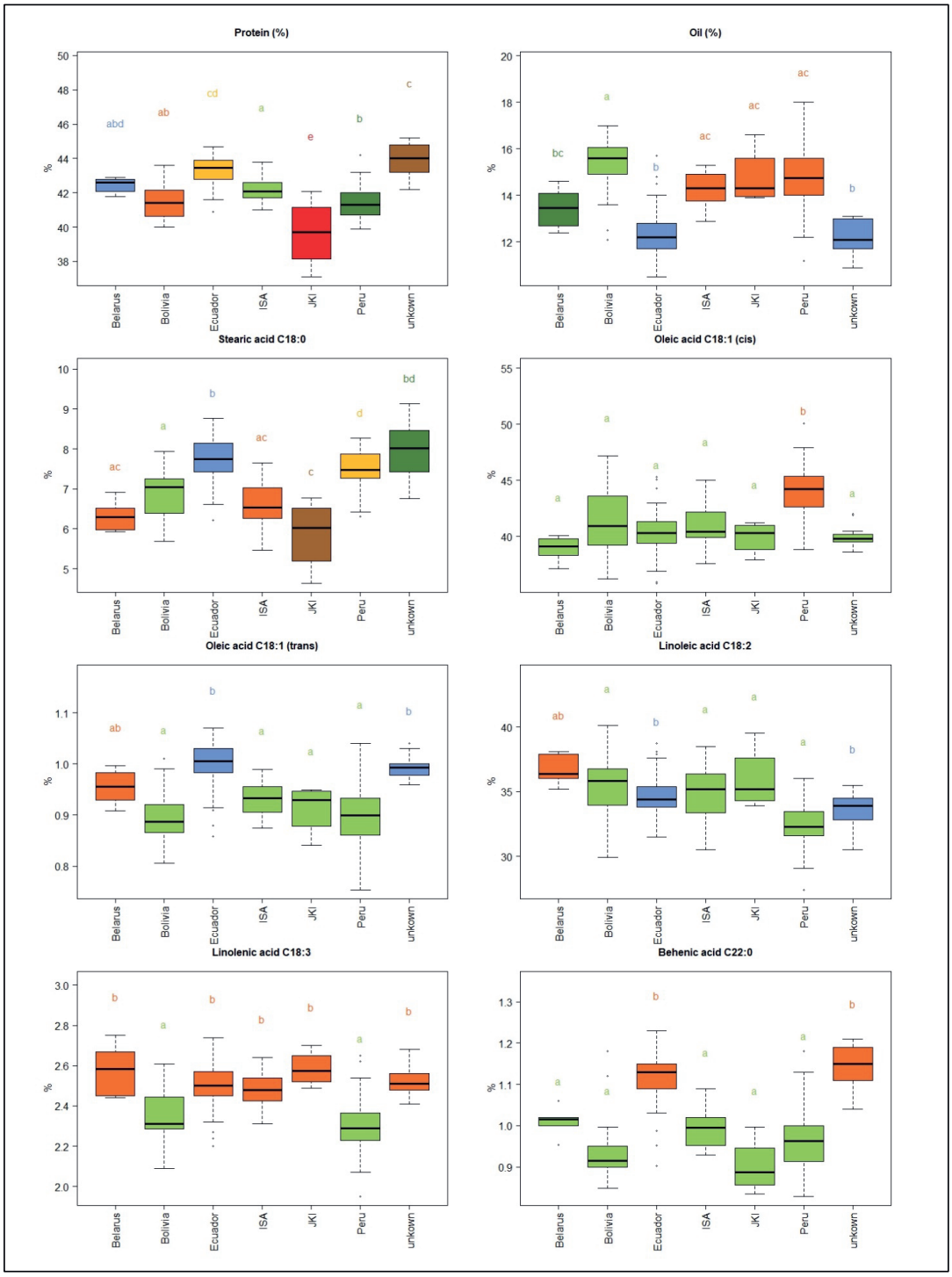


FIGURE 3| Boxplots depicting oil and protein concentration (% dm) in the seeds and fatty acids compositions in *L. mutabilis* seeds from accessions with different origins. Means marked with different letters are significantly different ($P < 0.05$), according to a Tukey post hoc test.

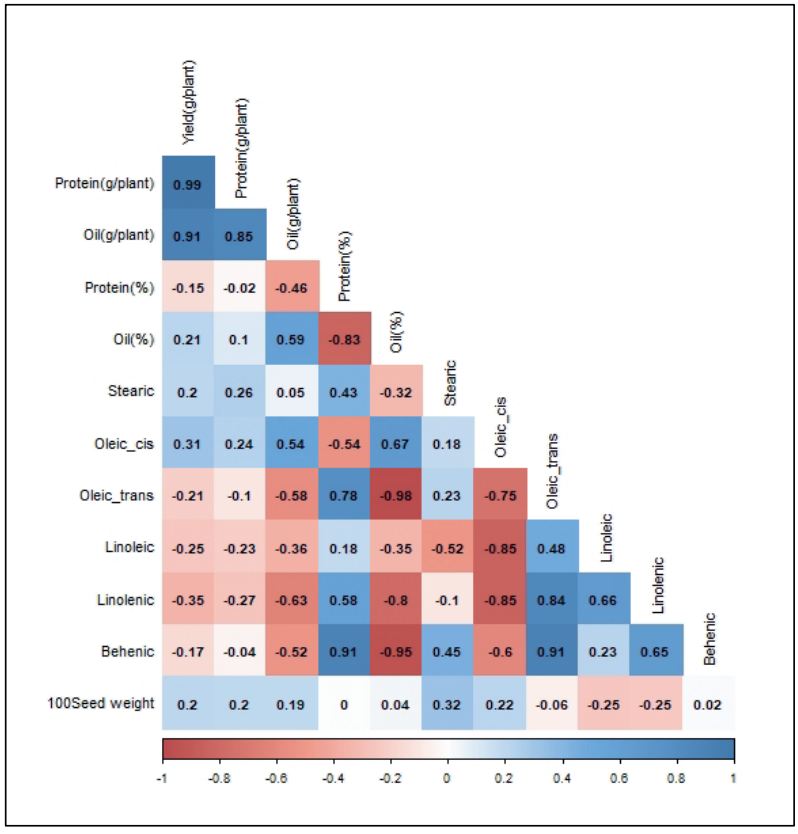


FIGURE 4| Correlation plot depicting the correlation values between predicted concentration of protein and oil (% dm) and fatty acids composition (% total FA area) in *L. mutabilis* seeds with seed, protein and oil yield (g/plant) and 100 seeds weight (g) for 220 accessions of Andean lupin.

Discussion

Breeding *L. mutabilis* as a multipurpose crop for protein and oil is crucial for the development of a sustainable alternative to soybean for Europe. The oil content of Andean lupin (~18%) is the highest within lupins and the one with a comparable quality to that of soybean (20%). However, unlike other plants, *L. mutabilis* is characterized by a recent domestication and a fragmented breeding history. Research and breeding efforts have long concentrated on small collections generated through mutations and inter-crossing, and only recently have started exploring the natural variability harboured by different germplasm pools. In the present study we characterize protein and oil content and oil composition for the first time in a collection of *L. mutabilis* accessions including both germplasm materials from the Andes and breeding lines from different institutes around the world. Herein we provide valuable information to drive the selection of

starting material for breeding programs aimed at the development of varieties for protein and oil production in Mediterranean conditions.

Among all the genotypes observed, a linear increase of both protein and oil yields (g/plant) with increasing seed yield was observed (r= 0.85; r= 0.91). This is not surprising because oil and protein are the major constituents of *L. mutabilis* seeds. However, no correlation was found between seed weight and concentration of oil and protein in the seeds, suggesting that breeding for larger and heavier seeds would not affect the concentration of protein and oil present in the seeds. On the other hand, the correlation between protein and oil concentration (%) in the seeds remains high and negative (r= -0.89). Correlations of oil and protein concentration are generally negative as the two components deposit simultaneously in developing seeds (Gomez *et al.*, 2011). Similar values of negative correlation between oil and protein content have been already reported in *L. mutabilis* (Clements *et al.*, 2008; Perez *et al.*, 1984) and are common for many oil crops (e.g. soybean, r=-0.79; (Kambhampati *et al.*, 2019); *Jatropha curcas*, r= -0.64; (Vaknin *et al.*, 2011)). From our analysis, emerges that germplasm material from Bolivia, Peru and JKI is characterized by higher oil concentration (14.7-15.2%), while material from Ecuador and unknown origin in the Andean region has overall a higher protein concentration (43-44%). These values are in line with the values reported for the Polish breeding material recently analysed by Czubinski et al. (Czubinski *et al.*, 2021) and characterized by 44.7 % of protein and 15.4 % of oil in the seeds. In addition, we also report the finding of outstanding accessions in our collection characterized by 45% of protein content in their seeds (LIB166, from INIAP collection) and 18 % of oil content (LIB126, from Peru). However, based on its highest seed yield (32.5 g/plant, ~3 t/ha), the accession LIB065 from Peru was characterized by the highest protein (13.2 g/plant) and oil (5.2 g/plant) yields, suggesting potential yields of 1.5 and 0.6 t/ha of protein and oil respectively, in selected varieties. Previous studies have reported even higher oil concentration for *L. mutabilis* seeds (up to 20 %), however it is important to keep into account that agronomic conditions may play a role next to the genetic component in determining this trait. Carvalho et al. (Carvalho *et al.*) reported a seed oil content twice as low in *L. mutabilis* genotypes under water stress. Similarly, the observed oil content in Mediterranean conditions might be influenced by the onset of drought during seed development, common of these environments. Noticeably, we report higher variability for oil content in the germplasm from Peru and Bolivia, which candidates these accessions as good starting material for breeding *L. mutabilis* varieties for oil production. It can be speculated that the adaptation of germplasm from Peru and Bolivia to water stress conditions in their native environments can contribute to the maintenance of high oil content in the seeds, even under drought stress in a Mediterranean environment.

Other than being characterized by an high oil content, *L. mutabilis* is also characterized by a fatty acid profile nutritionally superior to that of soybean: both have a similar ratio of saturated/unsaturated fatty acids (17–18%), but *L. mutabilis* has a lower amount of linolenic acid, thus avoiding the need for industrial removal of this fatty acid, as its oil stability is naturally higher (Schoeneberger *et al.*, 1982). In addition *L. mutabilis* oil does not have any toxic erucic acid, as

found in other lupin species, and when compared to other edible oils presents a higher or similar quality, being inferior only to olive oil (Martins *et al.*, 2016). In agreement with previous studies, we report an high share of oleic (C18:1(cis), 41.4 %), linoleic (C18:2, 34.1 %), palmitic (C16:0, 11.2 %) and stearic (C18:0, 7.4 %) acid in oil composition, accounting altogether for ~94 % of the total fatty acids in *L. mutabilis* seeds. Linolenic acid was present in very low amount, constituting only 2.4-2.7 % of oil content, well below the content of soybean comprised between 5.9-8.3 % (Sharma *et al.*, 2014). The fatty acids profile of *L. mutabilis* has been compared to the one of *L. albus* (Curti *et al.*, 2018), and of *L. angustifolius* (Czubinski *et al.*, 2021). The composition of *L. albus* is characterized by 54 % of oleic acid, 18.7 % of linoleic acid and 8.6 % of linolenic acid. *L. angustifolius* seeds have a higher amount of linoleic acid (40.3 %) and lower oleic (33.9 %) and linolenic acid (5.6 %) content. Comparing the results obtained for *L. mutabilis* accessions in our collection, we also indicate a more significant similarity to *L. angustifolius* fatty acid profile, with the advantage of a lower linolenic content in *L. mutabilis*, guaranteeing higher oil stability. Curti et al. (Curti *et al.* 2018), reported content of oleic acid up to 56% in *L. mutabilis* seeds from Argentina and Bolivia. In our collection, we report a lower oleic content for Bolivian material growing in Mediterranean conditions, ranging between 36.2-47.2 %. Instead, we report a higher oleic content in accessions from Peru, ranging from 38.8 to 50.1 % and the highest content in LIB126 (50 %), a genotype with white seeds, selected by the Peruvian Center for Research and Innovation in Andean Grains (CIINCA) which can represent, according to our results, a valid source of genetic variation to ensure a high oil production from *L. mutabilis* seeds. In our analysis, total oil content appears to positively correlate only with an increase in oleic acid percentage (r=0.6) while it negatively correlates with all the others fatty acids in our analysis. Previous studies assessing the effect of seasonal variation on oil crop's fatty acids profile have concluded that oleic content in seeds is highly influenced by temperature during seed development and to a lower extent correlated to total oil content (Demurin *et al.*, 2000; Fayyaz *et al.*, 2003). Fayyaz et al. (Fayyaz *et al.*, 2003) report that each increase of 1°C in temperature corresponds to 2 % increase in oleic content in sunflower seeds. Conversely, to an increase in oleic acid corresponds a decrease in linoleic acid, which appear as inversely proportional (r=-0.85). This close relationship between these two fatty acids is very common in oil crops and plays a vital role in the selection of high-oleic and high-linoleic varieties, depending on the desired properties of the resulting oil and the intended use of it.

The possibility of using *L. mutabilis* oil for the production of biofuel has been also considered (Zamora *et al.*, 2020). According to the similarity of the fatty acid profile of soybean (17-30% oleic acid, 48-59 % linoleic, 8-23.5 % palmitic and 2-5.4% stearic acid), the possibility of using *L. mutabilis* oil for the production of biodiesel seems promising. Furthermore a higher proportion of mono-unsaturated fatty acids is considered advantageous with regard to oil quality, which is the case for the higher proportion of oleic acid in *L. mutabilis* (40%) than soybean (17-30%). An high oleic content is desirable to improve both edible and industrial oil applications, as it facilitates improved health and oxidative stability for increased oil shelf life, flavour, durability and cold flow performance (Fayyaz and Ahmad, 2003).With regards to biofuel production, an

high amount of mono-unsaturated fatty acids leads to substantially less fuel polymerization than it would occur with fuel derived from polyunsaturated fatty acids, avoiding the formation of insoluble sediments and gums (Monyem *et al.*, 2001). As a rule of thumb, seed sources with higher amount of oleic, palmitic and stearic fatty acids are likely to produce significantly better quality biodiesel (Vaknin *et al.*, 2011).

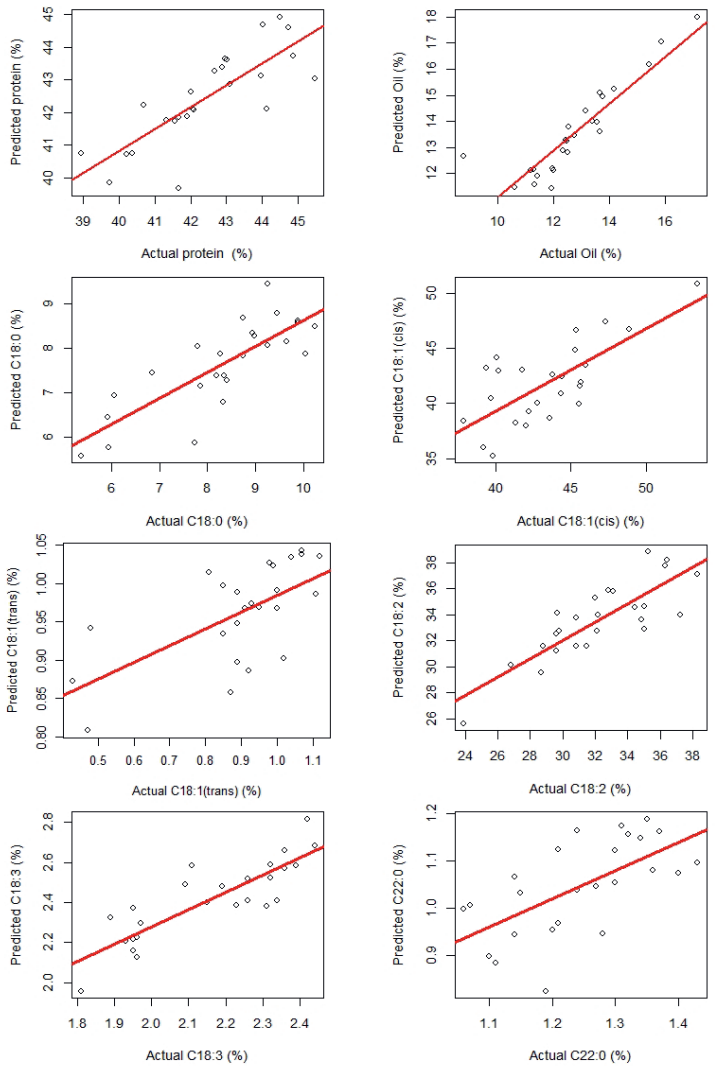
Thanks to its rich fatty acid composition, *L. mutabilis* oil can find applications both for nutritional and industrial purposes. Selecting for different fatty acids profiles is therefore important for breeding varieties for specific uses. The development of a high-throughput NIRS method to quickly characterize seed quality in large collections could be of great value in the selection of valuable breeding material. In this study we were able to create strong prediction models for protein and oil content (R^2 calibration=0.9, R^2 validation= 0.8) using a small subset of 28 samples, representing approximately 5% of the scanned material. Despite the fact that model accuracy will always rise with increasing sample size, as evidenced by the smaller difference between SEC and SECV in the model based on a larger calibration set (Table 2), our results show that valid predictions can also be obtained by selecting smaller calibration sets spanning a similar range of variation (Figure 2, PCA). These findings were further corroborated by the high linearity between the values predicted for protein content by the model based on 28 samples and the ones produced by a model developed using a calibration set of 94 samples (Figure 2, Correlation). Overall, the NIRS method was found to be highly predictive of seed protein concentration, oil concentration and linoleic, oleic (cis) and stearic acid proportions (R^2 = 0.9) and moderately good for behenic, linolenic and oleic acid proportions (R^2 = 0.7-0.9). Instead, it was not possible to accurately predict the ratios of palmitic, palmitoleic and arachidic acids. The reasons for the low-quality calibrations of some fatty acids include the low concentration and low variability of these components in *L. mutabilis* seeds, but could also be related to similarities in their NIR absorption pattern caused by the presence of the same absorbing molecular group (CH_2-) in different fatty acids (Sierra *et al.*, 2008). The inability to predict by NIRS fatty acids present in low ratios in the seeds of other oil crops has also been reported by other authors (Vaknin *et al.*, 2011). Still, considering that the primary objective of *L. mutabilis* breeding programs is currently the increase of oil content and oleic acid for high-quality oil production, our characterization by NIRS represents a good tool for selection. Furthermore, our characterization of 220 accessions from different origins, highlights the usefulness of using germplasm material from Bolivia and Peru as starting material for the development of *L. mutabilis* varieties characterized by high seed quality in Mediterranean winter cropping conditions.

Conclusions

Rapid and cost-effective analysis of *L. mutabilis* seed quality by NIRS is valuable for characterizing large collections of germplasm and identifying useful individuals for breeding applications. Our

findings indicate that accessions from different origins in our collection significantly differ in protein and oil content, but also in oil composition. The indications on outlying individuals can drive the selection of superior genotypes characterized by protein and oil content up to 18 and 45 % dm respectively, and an improvement of oil composition tailored to its end-use, e.g. by increasing oleic acid concentration for biodiesel production.

Supplementary Figure



SUPPLEMENTARY FIGURE 1| External validation of the NIRS calibration models in predicting laboratory measured values of protein content (% dm), oil content (% dm) and concentration of Stearic acid (C18:0), Oleic acid (cis and trans, C18:1), Linoleic acid (C18:2), Linolenic acid (C18:3) and Behenic acid (C18:22) in *L. mutabilis*

7

General Discussion

Introduction

The increasing need for a protein transition and a biobased economy in Europe has fueled the search for sustainable protein and biomass crops. This search becomes fundamental if we consider that the current agricultural scenario is characterized by the presence of few crop species or "major crops", relying on unsustainable, resource-intensive farming systems. In order to avoid dependence on few food crops and broaden crop diversity, there is a great need to boost the exploitation of plant genetic diversity. The importance of underutilized and neglected crops in a transition towards a more sustainable agricultural system has been internationally recognized, attracting the interest of researchers and breeding companies. Minor crops, known as such because they have so far played a minor role on the global market, can now represent valid alternatives to the current crops and serve as an excellent source of genetic resources also for the improvement of other commodity crops.

In this thesis we have focused on the potential of *Lupinus mutabilis*, an under-utilized Andean legume, as a multipurpose crop for Europe. The high quality of *L. mutabilis* seeds, characterized by a protein and oil content similar to soybean, has attracted increasing interest from Europe over the past decades. Moreover, its adaptation to low input farming in temperate climates candidates it as a superior alternative to the current plant-based source of protein and oil in Europe and other regions with temperate climates. Yet, despite the clear potential, research on *L. mutabilis* has been limited and its adaptation and cultivation has been restricted to its center of origin. The lack of genetic improvement and the narrow genetic diversity explored for important agronomic traits has hampered the development of this crop. To date, the vast majority of genetic resources for this species remain inadequately characterized and underutilized. In this study, we have extensively assessed -for the first time- the diversity and adaptation of a wide collection of *L. mutabilis* germplasm to European conditions, providing the phenotypic and genetic knowledge needed to drive the selection of locally adapted genotypes, and to implement molecular breeding tools that can speed up *L. mutabilis* breeding securing it as a promising crop for Europe. In the following sections, the insights generated in the previous chapters are comprehensively discussed and contextualized in light of the genetic improvement of *L. mutabilis* needed to establish this crop in Europe.

Exploiting germplasm pools as a source of genetic variation

The review of the history and past breeding achievement for *L. mutabilis* that we provide in **Chapter 2**, highlights a fragmented domestication and breeding history for this crop. A repeated use of small pools of genotypes, mainly generated by induced mutations or as a result of crossing between few genotypes, has characterized the few studies that assess the adaptation of this crop to Europe (Caligari *et al.*, 2000; Clements *et al.*, 2008; R. Galek *et al.*, 2017; Guilengue *et al.*, 2019; Galek *et al.*, 2016). The use of a limited amount of genetic resources has contributed to narrowing the genetic basis, which could explain the lack of adapted genotypes, characterized by early flowering and high yield in temperate climates, emerging from these studies (R. Galek *et al.*, 2017; Guilengue, Neves-Martins*, et al.*, 2020; Caligari *et al.*, 2000; Galek, *et al.*, 2016). In spite of the presence of considerable variation across germplasm collections of Andean lupin, these genetic resources have so far been unexploited for the breeding of *L. mutabilis* for Europe. The large size of germplasm collections, coupled with unavailability of detailed data and information and the often difficult accessibility of plant material, has often resulted in low use (<1%) of germplasm

resources not only in the improvement of *L. mutabilis*, but also of major crops as wheat (Dalrymple, 1986), maize (Dowswell *et al.*, 2019), soybean (Mikel *et al.*, 2010) and other grain legumes (S. Kumar *et al.*, 2004).

The genetic diversity of a species represents the most important reservoir of gene variants to draw on for the introduction of distinctive useful traits in crop adaptation and improvement. Hence, a significant effort should be devoted to broadening the genetic base of the material used in *L. mutabilis* breeding programs combining different strategies. The main strategy employed so far has been the use of induced mutation and intercrossing, which has not led to the successful introduction of a stable cultivar yet. An alternative strategy, is to use the genetically diverse germplasm from other geographical regions to cross with elite material selected (Qian *et al.*, 2007; Quijada *et al.*, 2004; Udall *et al.*, 2004). This approach requires a thorough characterization of the phenotypic and genetic diversity harbored in a germplasm collection (Qian *et al.*, 2007; Quijada *et al.*, 2004; Udall *et al.*, 2004) and optimally include an evaluation of the material across different agroclimatic areas that are of interest for the adaptation of the crop. Keeping into account these aspects, in this thesis we set up one of the largest collections of *L. mutabilis* ever assessed in Europe, including a wide share of Andean material and breeding lines from Europe, and evaluated its agronomic performance both as a winter crop for the Mediterranean region and as a summer crop for Northern/Central Europe (**Chapters 3 and 5**), but also investigated its seed and biomass quality through the development of prediction models for high-throughput phenotyping (**Chapter 4 and 6).**

Considering the limited availability of *L. mutabilis* lines resulting from breeding programs in Europe (**Chapter 2**), we set up a collaboration with the Instituto Nacional de Investigaciones Agropecunaria (INIAP) of Ecuador (Quito) in order to construct the first Unified EU collection of *L. mutabilis*. South American institutions currently host the majority of *L. mutabilis* germplasm, counting more than 3,000 Andean lupin genotypes, mainly conserved in the gene banks of Peru, Ecuador and Bolivia. INIAP collection of *L. mutabilis* includes approximately 370 accessions (Chalampuente-Flores *et al.*, 2021), representing probably the largest genetic diversity source for this species. On the basis of preliminary characterization of this collection, three geographically separated morphotypes have emerged as characterized by a considerable genetic and morphological variability and an ecological adaptation to different areas of the Andean region (Gross and Baer, 1981) (Tapia, 2015):

- Accessions from northern Peru and Ecuador, known as *lupine*, and characterized by prolific branching, very late flowering, great hairiness of leaves and stems, and tolerance to anthracnose;

- Accessions from central and southern Peru, known as *tarwi,* scarcely branched, moderately late and somewhat tolerant to anthracnose;

- Accessions from highlands of Peru and Bolivia, called commonly *tauri* and characterized by a smaller stature (1-1.4 m), a developed main stem, very early flowering and a high susceptibility to anthracnose

The inclusion of this diverse material in our assessment of diversity and agronomic performance of *L. mutabilis* is highly valuable, and it represents the first characterization of these morphotypes in European environments (**Chapter 3**).

Phenotypic and genetic assessment of the collection highlighted the presence of clear differences between the germplasm material gathered from the Andean region and the lines resulting from

breeding programs. When assessing germplasm material in Ecuador, no clear differences in phenotypic traits could be traced back to the different germplasm pools included in the collection. However, in Europe, a more defined clustering around germplasm pools of accessions with the same origin exhibiting similar phenotypes was noted (**Chapter 3, Fig. 3**). Accessions from the INIAP collection that were identified as coming from Ecuador or of unknown origin were characterized by high biomass yield, an indeterminate growth habit, poor seed yield and later flowering in European environments, but also by a high protein content in their seeds. Instead, accessions from Bolivia and Peru showed a certain proximity to breeding lines, were generally characterized by the presence of favourable agronomic traits (earlier flowering, higher seed yield, restricted branching), which led in certain cases to seed yields even superior to the one of lines bred for that specific environments, up to 3 t/ha and characterized by an high oil content and high concentration of oleic acid (**Chapter 3, Table 2, Chapter 6, Table 1 and 4**). Overall, from a phenotypic evaluation of the collection we concluded that landraces from Peru appear as promising starting material for breeding programs aimed at increasing seed yield in *L. mutabilis* varieties for European cropping conditions, while the use of Ecuadorian and Bolivian material remains highly valuable for the introduction of novel genetic variation into the breeding gene pool and for the selection of new genotypes adapted to Andean conditions. At the genetic level, the investigation of genomic relationships based on genotype information for each individual marker, highlighted the presence of two subgroups in the collection (**Chapter 4, Fig. 2B)**, confirming a main division of the accessions based on their origin and type (wild material vs. breeding material). One group includes most of the wild material from Ecuador, Peru, Bolivia and unknown origin, while the seconds group consist mainly of breeding lines and some promising accessions from Ecuador, Peru and Bolivia. Notwithstanding the presence of germplasm pools of high interest in our collection, grouping of the accessions was loose and revealed a low level of population structure and stratification in our collection, making this panel a good set for genetic association studies. Moreover, the analysis of principal components of phenotypic variation (**Chapter 3, Fig. 3**) highlights the presence of extreme phenotypes in our collections which can prove very valuable as starting material for breeding programs aimed at establishing selected accessions locally adapted to specific environments, or at designing a series of crosses between extreme genetic material to broaden the genetic basis of a breeding program.

The high variation present in the collection

Previous studies on *L. mutabilis* have explored only a small fraction of the natural genetic diversity of this species, using limited numbers of genotypes (12 to 30) in their evaluations of *L. mutabilis* (Caligari *et al.*, 2000; R. Galek, Kozak, Biela, Zalewski, Sienkiewicz*, et al.*, 2016; Clements *et al.*, 2008). The aim of this thesis was thus to broaden current knowledge on the extent of phenotypic variability characterizing *L. mutabilis* germplasm in terms of agronomic traits, adaptation and seed quality. Moreover, our study further widen the knowledge on this crop by investigating the biomass quality of agricultural residues to assess the potential of developing *L. mutabilis* into a multipurpose crop. This thesis is the first report on the evaluation of *L. mutabilis* as potential feedstock source for the European biorefinery industry, and that represents another novel aspect of our research.

The analysis of phenotypic variation in our collection of 225 *L. mutabilis* accessions revealed large ranges of phenotypic variability for 37 traits including agronomic traits (**Chapter 3, Table 2**), traits describing biomass quality and cell wall composition (**Chapter 4, Table 1**) seed protein and oil contents, and oil composition (**Chapter 6, Table 1 and 4**). A large range of variation in key

agronomic traits highlighted the presence of extreme phenotypes in the collection, ranging widely in plant height (24-208 cm), flowering time (65-127 days), biomass production (4.5 g- 2.1 kg), number of branching orders (0-4), seed weight (9.7-145 g/100 seeds) and production of pods and seeds on the main stem (0-26 pods, 0-68 seeds) and on the rest of the plant (0-49 pods, 0-108 seeds) (**Chapter 3**). Furthermore, the analysis of seed quality revealed the presence of accessions characterized by an oil concentration up to 18.3 % in seeds and a large range of protein content between 37 and 46.2 % (**Chapter 6**). As we expected, the variability reported in this study was superior to the one reported in previous studies of *L. mutabilis* diversity that used smaller sets of accessions (flowering time 65-88 days; seed weight 6.8-18.6; protein content 43.6-46.1 %; oil content 8-9.5 %) (Caligari *et al.*, 2000; Clements *et al.*, 2008; R. Galek, Kozak, Biela, Zalewski, Sienkiewicz*, et al.*, 2016). In addition, the variability of the present *L. mutabilis* collection was even superior to the variability reported for the global *L. albus* landrace genetic resources (Annicchiarico *et al.*, 2010), indicating that our panel is actually gathering a plethora of genetic diversity, comparable to global collections of genetic resources.

The use of this collection for the assessment of *L. mutabilis* potential as a source of lignocellulosic biomass, provides the first evidence that *L. mutabilis* cultivation in Europe can generate a consistent amount of biomass for the European biorefinery industry (**Chapter 4**). Based on our observation of biomass yields, we estimate a potential generation of 27 to 54 t/ha of agricultural residues when cultivating *L. mutabilis* in Mediterranean or North- European conditions, which could be valorized into a highly valuable product. To make this process economically feasible breeding for new accessions characterized by high biomass yield and high quality of biomass is key. Notably, we find that the biomass composition of these residues is comparable to the composition of other current sources of biomass, such as miscanthus and sorghum, and superior to the one of other lupin species (Hodgson *et al.*, 2010; Ţîţei, 2020; Y.L. Zhao *et al.*, 2009). The variability reported in the content of different fiber fractions and cell wall monosaccharide components was also comparable to the variability reported in other biomass crops with bioenergy applications, and coupled with a good heritability of the traits under study (H^2= 0.5-0.8) suggesting that breeding for contents of cellulose, lignin and pectic components (e.g., galacturonic acid and rhamnose) has a great potential to increase the genetic gain of *L. mutabilis* biomass quality.

Influence of different European environments on *L. mutabilis* agronomic traits

The environmental differences between the native environment of *L. mutabilis* and other cultivation areas around the world, such as Europe, represent one of the main challenges to the expansion of this crop. Due to a high susceptibility to low temperature (-2 °C) in the initial stage and a requirement of 350-800 mm of rainfall, sowing of *L. mutabilis* in Europe is limited to autumn in Mediterranean environments and to spring in northern countries.

The differences in photoperiod, temperature regimes and water availability related to the two different seasons and latitudes in which is possible to cultivate *L. mutabilis* in Europe, can highly affect phenotypic variation and adaptation behaviour of crops. Although fairly rare, the evaluation of germplasm collections of crops in different agroclimatic areas is particularly important to unravel the genetic and environmental influence affecting the expression of important agronomic traits under different conditions. These evaluations can be highly valuable for germplasm users, and reach both an international and possibly intercontinental target

audience (Annicchiarico *et al.*, 2010). In our analysis of agronomic adaptation to different environmental conditions, we noted a considerable influence of the environment on the morphology of *L. mutabilis*.

The effect of the environment was similar on most of the accessions but different across locations, as demonstrated by the low genotype per environment interactions (**Chapter 5, Table 1**), leading overall to the observation of two different morphotypes for *L. mutabilis* growing in Mediterranean winter cropping conditions and summer-North/Central European ones, which were different from the phenotype observed in Andean conditions (**Chapter 3**). In winter Mediterranean conditions, higher temperatures and water scarcity highly affected the architecture of all the accessions, limiting crop development to two branching orders and restricting biomass yield, but increasing seed yield overall in the panel. In North European summer cropping conditions, more constant and higher precipitations prompted *L. mutabilis* vegetative growth throughout the entire cropping season, resulting in highly branched plants holding floral buds and pods of no economic value at harvest. As also shown by these results, response to temperature and water availability can strongly influence the determinacy of the crop and as a consequence, seed yield. This effect is due to the characteristic indeterminate growth habit of this plant which leads to preferential partitioning of nutrients towards vegetative growth (Hardy *et al.*, 1998) in the presence of water availability and lower temperature.

Another important factor is photoperiod, which together with temperature, is known to trigger the beginning of flowering in many species, including lupins. Taking into account that the development of early flowering lines has been highly recommended in order to grant *L. mutabilis* adaptation to European environments, a clarification of the mechanism underlying flowering time regulation in this crop is urgently needed. Flowering time in lupin species has been shown to be highly responsive to prolonged exposures to cold (also known as vernalization period), ambient temperatures and photoperiod (Kedar N. Adhikari *et al.*, 2012; Reader *et al.*, 1995). Previous reports on the photoperiodic sensitivity of *L. mutabilis* are contrasting (Hackbarth, 1961; Jacobsen and Mujica, 2008), but flowering time has been reported as highly responsive to the environment, with a predominant effect of temperature (A. Hardy *et al.*, 1998). Our results suggest a combined effect of photoperiod and temperature on *L. mutabilis*. Short days in Portugal did not significantly affect flowering time (110 days), which was in line with the general flowering time observed in Ecuador under a fixed photoperiod. Conversely, long days in Netherlands remarkably shortened flowering time by about a month (to 80–90 days). However, in both European environments, it was possible to distinguish earlier and late flowering accessions characterized by a different growing degree day (GDD) accumulation. In Portugal, early accessions had a GDD of about 600 °C and flowered before 105 days after sowing, while late accessions flowered between 116 and 127 days with a GDD of about 1,000 °C. In The Netherlands, some accessions were very early and flowered between 73 and 80 days after sowing (GDD 700 °C), while a higher proportion was late and flowered between 90 and 104 days (GDD 1,200 °C).

Effect of extreme climates in the same locations, also brought significant differences in the adaptive responses of *L. mutabilis* to two consecutive growing years in The Netherlands. Climate change plays a big effective role in this, considering that 2019 was the first warmest year on record in The Netherlands since 75 years. Moreover, 2019 was characterized by an extreme scarcity of rainfall during summer (2.25 times less than in 2020). Studies on the effect of rainfall on the development and yield of *L. angustifolius* in temperate climates have shown that seed yield is dependent to the largest extent on the amount of rainfall in June and July, which are the periods of blooming and pod setting (Podlesny and Podlesna, 2011). Accordingly, we observe significant

decrease in seed yield in 2019 in The Netherlands, to almost one third of the seed yield registered in 2020 (**Chapter 3**). Consequently, a similar decrease is also observed on the fresh biomass produced, the number of branching orders developed, seeds and pods produced and seed weight, indicating a fundamental role of water availability during the delicate phase of blooming and pod setting. Hence, the integration of additional irrigation under such climatic extremes appears fundamental in these developmental phases to ensure a good yield in *L. mutabilis*. Furthermore, the order of the accessions differed in these two locations indicating that different accessions are characterized by different responses to environmental stimuli, that under climatic extremes, can be useful to select for. The large extent of environment-specific responses and the partly different adaptive traits underlying these responses support the development of distinct varieties for contrasting agroclimatic regions and climatic extremes. Hence, we decided to always analyse these two locations in The Netherlands (NL-Sc and NL-Wi) separately.

The environmental component of the phenotypic variation can be understood as an adaptive behaviour of *L. mutabilis* development under specific environmental factors, which is independent of the heritable fraction of the phenotypic variation. Hence, despite the larger heritability of agronomic factors such as plant height (0.8), flowering time (0.9) and number of pods and seeds on the main stem (0.6) across environments, environmental factors influencing *L. mutabilis* architecture will have strong implications for breeding, as also demonstrated by our identification of QTL for specific environments (**Chapter 4, Chapter 5**). In particular, the different development of *L. mutabilis* growth in response to the different cropping conditions, suggests also the possibility to develop varieties for seed yield in Mediterranean conditions, and as a source of lignocellulosic biomass in Northern-European conditions.

Breeding for seed quality and biomass quality: can we breed for both simultaneously?

The introduction of new crops on the market should call for a careful assessment of their potential as multipurpose crops. Taking into account the time and investment required for the introduction of a new crop on the global market, the advantages of these species in comparison to the major crops already existing on the market should be clear and in line with the current needs of our agricultural system. New crops should be successfully integrated with farming to improve crop yields, diversify products and improve farm-sustainability in the long term. Crop species designed to optimize the use of different plant parts for the production of food, energy, feed and by-products are now in high demand since they can guarantee a better resource use-efficiency in crops, reducing waste and maximizing outputs through a full utilization of the plant.

The scope of LIBBIO, a European union founded project, was exactly to develop Andean Lupin varieties adapted to European farming conditions that are optimized for the production of food and non-food products. Within the framework of LIBBIO, we explored the quality and variation of biomass composition in both Mediterranean and North-Central European conditions, reporting that a conversion of *L. mutabilis* biomass into fermentable sugars for bioenergy production appears possible and promising in both conditions (cellulose yields up to 500g/kg DM, 54 % of glucose in the cell wall, **Table 1 , Chapter 4**). However, due to the high environmental effect noted in those contrasting environments in Europe, it would be possible to foresee a first introduction of *L. mutabilis* for different end-uses in different countries, while working on the development of advanced varieties that can combine an optimal production of both seed protein and oil, and biomass for the biorefinery industry. At the moment, based on the

striking difference of biomass production and seed yield on different environment, it would be advisable to start the introduction of *L. mutabilis* in Europe as a protein and oil crop in the Mediterranean region and as a biomass crop in North-Europe. Such an approach would have the advantage of establishing in parallel new processing technologies for the extraction and production of high-value products, giving Europe both innovative potential and a competitive advantage as the leader in developing new Andean lupin applications and markets.

Notwithstanding the high nutritional value of *L. mutabilis* oil, the similarity with soybean in oil content and fatty acid profile (18 % oil content, 17-30% oleic acid, 48-59 % linoleic, 8-23.5% palmitic and 2-5.4% stearic acid), candidate this crop also as a possible source of oil for biofuel production. In **Chapter 6**, we report the characterization of seed quality and oil composition for *L. mutabilis* growing in Mediterranean conditions, where seed yield was higher. The higher proportion of mono-unsaturated fatty acids in *L. mutabilis* oil (40%) than in soybean (17-30%) represents a further advantage for this crop, as a higher oleic content is desirable to improve both edible and industrial oil applications since it facilitates improved health and oxidative stability for increased oil shelf life, flavor, durability and cold flow performance (Fayyaz *et al.*, 2003) Furthermore, our indications on the range of variation in fatty acids profiles across different germplasm pools can be used to design crossing schemes leading to higher yield of oleic, palmitic and stearic fatty acids, which are likely to produce significantly better quality biodiesel (Vaknin *et al.*, 2011).

Mono-directional breeding towards a single products can focus on maximizing the yield of a single component, hence being more straightforward to achieve. Breeding for high-value by-products and full utilization of plants, requires a more thorough understanding of the interactions between the different traits, such as biomass production and seed yield, especially at the genetic level. Due to the scarcity of genetic knowledge on this species, more work (and time) is needed to fill the gap in *L. mutabilis* breeding and to achieve a full exploitation of this species through its development as a multipurpose crop. At present, the development of stable lines adapted to specific conditions and targeting as end products either biomass or seed production appears as the most feasible option to grant a rapid introduction of *L. mutabilis* on the European market.

A first step towards the implementation of marker-assisted selection in *L. mutabilis*

The absence of molecular and genetic tools in neglected and underutilized species, currently represents one of the main hindrance to the development of better genotypes/varieties. The need to accelerate the application of multi-omics approach to exploit their potential and address current agricultural and environmental challenges, has been widely recognized by the scientific community. Conventional breeding relies on the generation of varietal parents through mass selection and outcrossing, and the development of cultivars through inbreeding and hybrid breeding (Salentijn *et al.*, 2015). This approach is based on phenotyping of many plants for several generations, a costly and time consuming process, which becomes prohibitive for breeding companies and for a fast introduction of new varieties into the market. Conversely, the use of molecular markers as breeding tools, can assist and speed up breeding and reduce the need to phenotype large number of samples.

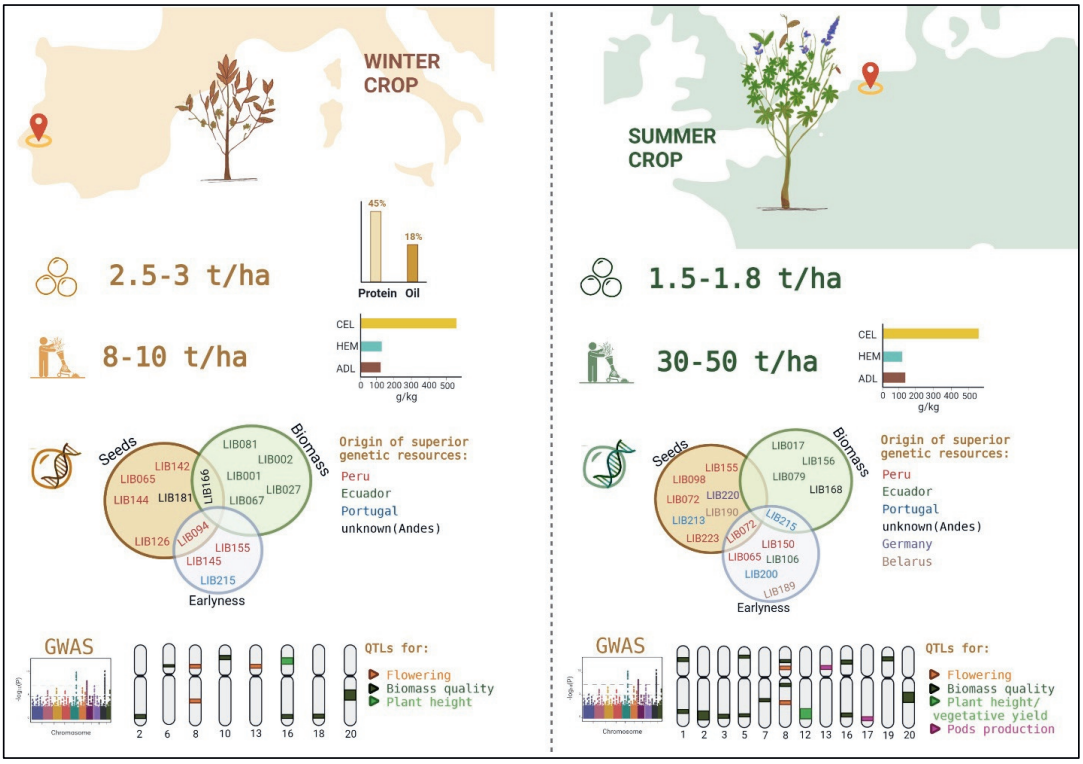


FIGURE 1| Schematic representation of the findings of this thesis. From our study, *L. mutabilis* emerges as a promising crop for seed production in the Mediterranean region (2.5-3 t/ha) and as a competitive source of lignocellulosic biomass for North-Central Europe (30-50 t/ha). The seeds represent a source of protein and oil, while the vegetative residues a good source of cellulose. The extensive phenotypic and genotypic characterization of a large panel of accession here conducted, provides a selection of superior genotypes characterized by important traits related to seed quality, early flowering and biomass production and a set of molecular tools (QTL) to aid the breeding of this crop.

Markers development for marker assisted selection schemes has not yet been implemented in *L. mutabilis*. The majority of molecular studies on this species have so far focused on understanding *L. mutabilis* phylogeny or used old-generation markers (RFLP, AFLP, ISSR and RAPD) developed for other lupin species to assess diversity within *L. mutabilis*. However, new developments in genomic technologies now provide a realistic opportunity to overcome the scarcity of genomic information and to hasten the identification of traits of interest. Many important agronomic traits are quantitative, thus determined by polygenic effects i.e., the product of two or more genes, and their environment. Hence it is often possible for agronomic traits to associate genetic regions to phenotypic traits, through the identification of quantitative trait loci (QTL). In the past, QTL development has mainly been based on mapping of biparental populations, limiting the detection of important alleles to the differences between two contrasting genotypes and few recombination events. However, thanks to the recent advances in sequencing technologies it is now possible to identify QTL on large panel of accessions, by generating thousands of markers at a much lower cost aiming at capturing all genetic diversity and associating it to complex agronomic traits, through a genome wide association approach (GWAS). GWAS overcome the limitations of biparental crossing and takes advantage of the historical recombination events that have accumulated over thousands of generations (Korte *et*

al., 2013), eliminating the need to invest multiple years in creating several recombinant generations. Hence, this approach appeared as the most suitable to exploit the genetic diversity harbored in *L. mutabilis* germplasm collections, and we selected it for our study.

A genome-wide approach

The lack of genetic knowledge and of a genome sequence for *L. mutabilis* did not provide an optimal basis for GWAS at the start of this thesis. While waiting for the release of a genome assembly for *L. mutabilis,* we decided to use the recently published *L. angustifolius* genome (951 Mb; 2n =40), as reference for mapping our sequences (Hane *et al.* 2017). Legumes are characterized by a substantial genome synteny across the entire family, therefore conservation of homologous genes and gene order between genomes of different legume species is expected (Jha *et al.*, 2022; Ren *et al.*, 2019). Furthermore, similar 2C nuclear DNA contents were estimated in *L. mutabilis* (1.90 pg) and *L. angustifolius* (1.89 pg), suggesting that there might be a higher affinity between these two species, notwithstanding the difference in number of chromosomes (Naganowska *et al.*, 2003). An average mapping rate of 76.9% and a properly paired average of 63.7%, confirmed a quite good alignment of *L. mutabilis* SNPs to *L. angustifolius* genome, if we consider the difference in species and the use of a still fragmented genome as reference (**Chapter 4**). This method enabled us to develop a prototype of a genome-wide association analysis for *L. mutabilis,* which gives a blueprint of the genetic architecture of agronomic traits and biomass quality in this species.

In the present study, we identified a total of 71 QTL associated to agronomic traits and biomass quality in *L. mutabilis* (**Chapter 4, Chapter 5**). These QTL provide the first insight on the genetic architecture of important breeding traits which are relevant for the development of *L. mutabilis* cultivars adapted to European environments and targeting different end-products. We reported the finding of relevant QTL on chromosome 8 and 13 linked to high variation in flowering time (11-16%), which could be used for the development of early flowering varieties in different environments. Furthermore, our analysis of association of genomic regions with phenotypic traits confirms the high effect of the environments on *L. mutabilis* development, leading to the identification of QTL for specific environments. The presence of a major QTL for vegetative yield was found in North-European cropping conditions, where biomass production was very advantaged, and could be used to stabilize high yield of biomass production. The combined use of this QTL and QTL reported for biomass quality (**Chapter 4**) can directly improve the quantity and quality of biomass produced, making possible the use of this crop as a feedstock source for e.g., bioenergy production. The identification of QTL linked to strong reduction in lignin and pectin content, and increase in cellulose content can lead to significant improvement in biomass quality, as it has been the case also for other crops such as maize or miscanthus. The candidate genes identified provide useful insights into the molecular mechanism affecting plant development and cell wall composition in *L. mutabilis*, which have never been investigated previously. Furthermore, the finding of genes already used to enhance biomass quality in other crops, e.g. through increase in cellulose content (*CESA4* in shrub willow; Serapiglia et al., 2012); regulation of monolignol biosynthesis (*SKIP31* in sugarcane; Hodgson-Kratky *et al.,* 2021), reduction of lignin content (*CCR-SNL6* in rice; Bart *et al.*, 2010) or improved cell wall hydrolysis (*RAV1* in sugarcane; Tavares *et al.,* 2019) appears promising for a tailored improvement of *L. mutabilis* biomass quality for biobased applications. Of particular interest, is the finding of this *CESA4 (LOC109334350)* which appears to be syntenically conserved only among Fabaceae and to be homologous to another *CESA4* conserved across all angiosperms (Pancaldi, unpublished results). Homologous genes can diverge in functionality when are pushed to different genomic contexts across different families by evolution (Zhao *et al.*, 2017; Kerstens *et al.*, 2020), leading

often to a neo-functionalization of the gene representing an evolutionary advantage, which in this case might specifically affect cellulose/cell wall in Fabaceae.

The identification of QTL for agronomic traits in *L. mutabilis*, reflected a high degree of interconnection between growth habit, stem length, biomass production, flowering time and yield, characteristic of many undomesticated grain legumes (Krylova *et al.*, 2020). Determinate growth habits have been achieved in many legumes through the mutation of single genes (i.e., *Terminal Flower 1* in pea, fava bean, common bean and soybean), while traits as flowering time have been described as quantitative and controlled by major QTL and numerous regulatory genes, including for e.g. in *L. albus FLOWERING LOCUS T (Fta1), CONSTANS (CO), FY, MOTHER OF FT AND TERMINAL FLOWERING LOCUS 1 (TFL1),* interacting with genes related to photoperiod, vernalization and gibberellin pathway (Rychel *et al.*, 2019). In our panel the identification of different regulators of flowering time across environments linked to the hormonal pathway (*GAMYB, BIM1*) and the circadian regulation (LNK) and acting upstream of genes like *LFY*, *FLOWERING LOCUS C* (*FLC*), *CO* and *FT*, indicates that the same genes described for legumes must play a key role also in controlling flowering in *L. mutabilis,* in agreement with early reports for other lupin species (Nelson *et al.*, 2017; Rychel *et al.*, 2019; Taylor *et al.*, 2019). Additionally, we ascribe differences in vegetative development across European environments to two different QTL, respectively on chromosome 12 for North-European environments and on chromosome 16 for Mediterranean ones, harboring different transcription factors involved in hormonal pathways (i.e., *GA30X1* in the gibberellin pathway) and responsive to environmental stimuli (i.e., *Dr1* and *HDA15* such as drought and elevated temperatures). These results suggest that studying a biological process under diverse environments captures a more complete picture of its regulatory pathways, and can therefore lead to more valuable insights for breeding for locally adapted cultivars and shed lights on the different mechanisms regulating environmental-responses in this crop.

Future steps in the development of *L. mutabilis* as a crop for Europe

The plethora of data and information gathered throughout this thesis, provides the framework for a much wider investigation and deeper understanding of *L. mutabilis* potential. We focused on delivering a phenotypic and genetic screening of all *L. mutabilis* germplasm resources we were able to gather, laying the foundation for an accurate selection of valuable breeding material and for the initiation of large breeding programs of *L. mutabilis* in Europe. Breeding of locally adapted genotypes for Mediterranean and North-Central European cropping conditions can be assisted and speed up by the development of molecular markers on the basis of the knowledge generated in this study. As to that, our identification of QTL and candidate genes for agronomic traits and biomass quality appears promising and could be further validated upon the release of *L. mutabilis* genome.

Selection of superior genotypes for different European environments

The work reported herein characterizes the largest diversity panel of *L. mutabilis* accessions studied so far, paving the way to an informed selection of valuable lines for future breeding. The identification of several accessions characterized by relevant breeding traits could now aid the introgression of desirable traits for specific environments into stable lines, making it possible to develop *L. mutabilis* into a valuable crop for Europe. The selection of appropriate genotypes to serve as parents in crosses is one of the most important decisions faced by plant breeders, that will facilitate the exploitation of maximum genetic variability and the production of superior

recombinant genotypes (Bertan *et al.*, 2007). Although many advances in biotechnology and bioinformatics tools have been made, it is still common for the breeder to select parents based on their phenotypic performance regarding specific characteristics. However, thanks to a thorough characterization of natural diversity and to the development of molecular markers as part of this thesis, it will now be possible to guide more precisely the introgression of desirable traits from donor individuals into stable lines. As extensively discussed in **Chapter 2**, *L. mutabilis* introduction in Europe relies on the selection of genotypes characterized by determinate growth, early flowering and high seed yield. Other than superior individual performance only, parental selection for crosses should also take into account the ability of genotypes to positively react to specific environmental stimuli, hence to adapt to particular environmental conditions. The presence of large variation in agronomic traits, biomass quality and seed quality observed in the collection under study allowed the identification of potential superior genotypes for different traits. Furthermore, due to a major effect of the environment on agronomic traits, in different locations and under different climatic conditions we identify different accessions characterized by a superior performance (**Chapter 3, Table 3 and 4**). To sum up our results and facilitate the selection of superior lines for future breeding programs, we have schematized our findings in Fig. 1 and gathered our observation about agronomic traits (**Chapter 2**), biomass quality (**Chapter 3**) and seed quality (**Chapter 5**) characterizing these lines in Table 1 and 2 of the Annex. These results provides a blueprint to guide breeders in their selection of parental lines, to then cross them according to selected crossing schemes. Genotypes characterized by different agronomic traits -which would be desirable to combine in one variety- could be selected and used for the development of new varieties e.g., through repeated rounds of backcrossing and generation of Near Isogenic Lines (NILs) or through hybrid breeding, to harness heterosis. However, some of these approaches require the selection of uniform and stable genotypes to start with. Due to the considerable variation still noted within accessions, uniformity should first be granted through rounds of self-crossing, especially for the germplasm resources gathered from the Andean region. Different ranges of cross-fertilization by insects have been reported in *L. mutabilis*, fluctuating from 4–11% in Peru to 9.5–18.9% in Poland (Gnatowska *et al.*, 2000; Blanco Galdos, 1982). The isolation of different genotypes is thus highly recommended in breeding programs, as much as the careful wrapping of emasculated flowers in intraspecific hybridization (Adomas *et al.*, 2015; E Von Baer, 2011).

Additionally, the selection of highly diverse parental lines can be exploited to further broaden the genetic base of the crop, by creating a breeding population of *L. mutabilis* for Europe where lines with very different genetic backgrounds are crossed to generate additional genetic variability in the progeny. Cycles of recombination and selection aimed at accumulating a high proportion of desirable alleles on individual plants, can maximize overall genetic improvement. In this process, the use of molecular markers and high-throughput phenotyping methods can significantly reduce time and costs associated with breeding. In **Chapter 4** and **6** we have already demonstrated how the use of Near Infrared Spectroscopy can contribute to a high-throughput phenotyping of biomass and seed quality, and developed accurate prediction models that can be used to further reduce the cost and time associated with phenotyping. Once superior progenies are selected, working on a reduced number of accessions will highly increase the economic feasibility of having multi-year and multi-environment trials, leading to more accurate estimations of genotype performance, before proceeding to the selection of locally adapted varieties.

Exploiting natural allelic variation through genomic assisted breeding

Over the past 15 years, genomic assisted breeding has expedited timelines of breeding progress across a broad range of crop species, with the development of more than 130 publicly bred cultivars of different crops (Vogel, 2014). This advancement has been possible only thanks to a remarkable rise in throughput and accuracy of genome sequencing technologies. High-density genotyping systems, in combination with high-throughput phenotyping platforms, enable the survey of comprehensive diversity panels available in many crops to associate the genome information with variations in important phenotypes (Varshney *et al.*, 2021). Genotyping-by-sequencing of *L. mutabilis* collection in this study generates the first reduced representation of the genome of this species. However, mapping of these sequences on *L. angustifolius* genome, in the absence of a *L. mutabilis* genome, comes at the cost of a certain degree of uncertainty in the localization of QTL and candidate genes. The individuation of candidate genes in a window-region identified by the extent of linkage disequilibrium, assumes a full conservation of homologues genes and of their order between these two species, however studies assessing synteny of the two genomes are not available yet. A low degree of synteny, due for example to chromosome rearrangements, could explain the missed identification of major genes already identified in *L. angustifolius*. On the other hand, key agronomic traits could be also controlled by different regions *in L. angustifolius* and *L. mutabilis*, as already shown for example for *L. albus* and *L. angustifolius* (Książkiewicz *et al.*, 2017). Therefore, the best approach would be to re-align the generated sequencing information to *L. mutabilis* genome, once assembled and available, to generate a new set of SNPs. The combination of these SNPs and extra phenotypic data, gathered from carrying out these field trials for more years in the same locations, would certainly pave the way to the identification of relevant associations of allelic variation with phenotypes. Furthermore, genome-wide markers can also be used for genomic selection (GS), to estimate breeding value on the basis of the effects of all loci. Methods like GS can facilitate the assessment of genetic worth of the vast genetic resources archived in germplasm collections and guide the genetic exchange in breeding programs aimed at enhancing crop germplasm.

Annex

TABLE 1| Accessions selected as superior for cultivation of *L. mutabilis* as winter crop in Mediterranean environment, based on the different traits evaluated.

	Genotype	Origin	Seed yield	Vegetative yield	Flowering time	Branch orders	Protein content	Oil content	CEL g/kg	HEM g/kg	ADL g/kg	Glucose % CW
Seed yield & quality	LIB065	Peru	**32.4**	50.3	102	1.8	43.21	15.91	520.00	134.00	154.00	49.32
	LIB142	Peru	**26.7**	53.7	107	1.5	41.67	16.50	503.33	129.33	133.33	47.88
	LIB144	Peru	**26.6**	33.7	105	1.5	42.20	15.46	509.33	130.67	138.00	50.71
	LIB181	Unknown	**26.3**	62.6	115	2	45.18	13.13	526.00	132.00	129.00	50.84
	LIB094	Peru	**26.1**	41.2	**97**	1.3	41.32	16.08	523.33	131.33	136.00	51.67
	LIB166	unknown	17.93	62.84	115	1.9	**45.15**	11.49	51.49	528.00	138.67	123.33
	LIB126	Peru	21.5	28	104	1.1	40.00	**17.96**	49.62	520.67	124.67	122.67
Biomass yield & quality	LIB081	Ecuador	10.8	**74.2**	125	1.6	**46.05**	11.03	518	130	126	**51.78**
	LIB002	Ecuador	22.67	**89.4**	115	2	44.82	13.36	**538**	122	128	**52.08**
	LIB001	Ecuador	21.5	57.7	114	1.8	45.94	12.08	**522.7**	126	131.3	46.7
	LIB027	Ecuador	15	50.2	115	1.5	44.01	12.01	**508.7**	137.3	124	48.8
	LIB067	Ecuador	16.8	28.3	101	1.3	42.08	14.82	**505.3**	108.7	126	48.3
Earliness	LIB155	Peru	17.7	40.9	**92**	1.2	41.51	16.29	518.00	137.33	148.00	50.50
	LIB145	Peru	20.3	29.3	**94**	1.2	41.51	16.29	508.00	120.00	128.00	49.58
	LIB215	Portugal	12.7	27.0	**96**	1.3	42.00	14.37	503.33	132.00	138.67	47.91

TABLE 2| Accessions selected as superior for cultivation of *L. mutabilis* as summer crop in North European environments, based on the different traits evaluated. Due to the large differences between NL-Sc and NL-Wi trials, we report both their averaged means and the data relative to each of the trial separately.

	Genotype	Origin	Seed yield	Vegetative yield	Flowering time	Branching orders	CEL g/kg	HEM g/kg	ADL g/kg	Glucose % CW
Seed yield	**LIB155**	**Peru**	9.38	153.38	78.38	2.64	486.00	129.60	126.40	52.46
		NL-Sc	2.21	136.58	89.26	1.76	476.67	126.67	123.33	47.56
		NL-Wi	16.55	170.18	67.50	3.52	500.00	134.00	131.00	59.80
	LIB098	**Peru**	8.27	137.73	89.49	2.32	507.00	129.67	130.33	54.13
		NL-Sc	1.29	111.92	100.14	2.06	498.67	128.67	125.33	50.18
		NL-Wi	15.26	163.53	78.84	2.58	515.33	130.67	135.33	58.09
	LIB220	**Germany**	7.64	114.46	78.27	2.10	457.00	136.50	128.50	44.90
		NL-Sc	1.09	102.27	85.78	2.31	456.67	140.00	128.67	43.11
		NL-Wi	14.19	126.65	70.77	1.90	458.00	126.00	128.00	50.26
	LIB213	**Portugal**	9.28	88.17	73.92	2.85	492.40	123.60	128.40	51.04
		NL-Sc	4.38	76.10	80.22	2.73	492.00	127.33	126.67	48.59
		NL-Wi	14.19	100.24	67.62	2.97	493.00	118.00	131.00	54.70
	LIB190	**Belarus**	7.51	199.64	79.36	2.25	494.80	126.00	126.80	52.92
		NL-Sc	1.51	277.80	90.01	1.93	462.00	129.00	119.00	45.62
		NL-Wi	13.51	121.49	68.70	2.58	516.67	124.00	132.00	60.23
	LIB072	**Peru**	7.02	69.59	77.30	1.38	479.00	111.00	115.00	53.59
		NL-Sc	11.26	100.80	88.62	1.28	480.00	114.00	116.00	46.72
		NL-Wi	2.79	38.38	65.99	1.48	478.00	108.00	114.00	57.03
	LIB223	**Peru**	4.66	289.76	94.21	2.48	510.00	110.50	117.50	56.67
		NL-Sc	7.53	434.22	95.80	2.42	502.67	108.00	111.33	53.70
		NL-Wi	1.78	145.30	92.62	2.54	532.00	118.00	136.00	65.59
Biomass yield & quality	**LIB017**	**Ecuador**	4.90	248.20	89.37	1.89	516.00	119.67	120.67	60.00
		NL-Sc	4.40	129.07	91.47	1.47	507.33	121.33	115.33	53.54
		NL-Wi	5.40	367.32	87.27	2.31	524.67	118.00	126.00	66.46
	LIB156	**Ecuador**	3.65	274.89	87.31	2.18	521.00	124.33	129.00	58.31
		NL-Sc	1.89	126.37	85.69	1.46	516.67	128.00	126.00	53.61
		NL-Wi	5.41	423.41	88.92	2.89	525.33	120.67	132.00	63.01
	LIB079	**Ecuador**	3.42	361.76	99.53	2.39	516.00	117.33	119.00	57.89
		NL-Sc	3.20	489.61	103.04	1.96	499.33	118.00	113.33	51.50
		NL-Wi	3.64	233.92	96.01	2.81	532.67	116.67	124.67	64.28

	Genotype	Origin	Seed yield	Vegetative yield	Flowering time	Branching orders	CEL g/kg	HEM g/kg	ADL g/kg	Glucose % CW
Biomass yield & quality	**LIB168**	**unknown**	4.57	304.36	91.58	2.09	521.33	134.00	123.00	58.03
		NL-Sc	2.45	424.42	91.53	1.80	518.67	138.00	116.67	54.71
		NL-Wi	6.70	184.30	91.63	2.38	524.00	130.00	129.33	61.34
	LIB215	**Portugal**	3.47	102.38	76.45	2.35	469.67	118.67	128.33	51.78
		NL-Sc	0.80	147.32	85.21	2.29	470.00	124.00	125.33	48.14
		NL-Wi	6.14	57.43	67.69	2.41	469.33	113.33	131.33	55.43
Earliness	**LIB200**	**Portugal**	6.44	74.04	72.84	2.89	491.50	126.67	129.17	52.64
		NL-Sc	1.40	74.04	79.92	2.23	484.00	127.33	131.33	48.61
		NL-Wi	11.48	-	65.75	3.56	499.00	126.00	127.00	56.68
	LIB065	**Peru**	6.33	85.69	80.41	2.36	506.67	120.00	122.33	57.67
		NL-Sc	3.49	112.96	94.70	2.04	486.67	118.00	116.67	51.26
		NL-Wi	9.17	58.41	66.12	2.69	526.67	122.00	128.00	64.07
	LIB189	**Belarus**	3.61	95.44	79.64	2.67	488.67	127.67	121.00	54.41
		NL-Sc	0.96	109.16	76.09	2.48	467.33	131.33	117.33	47.59
		NL-Wi	6.25	81.72	83.19	2.85	510.00	124.00	124.67	61.22
	LIB150	**Peru**	4.06	107.95	72.36	2.20	512.17	118.50	124.67	56.67
		NL-Sc	0.79	107.95	75.62	1.68	497.33	132.00	123.33	51.64
		NL-Wi	7.32	-	69.09	2.71	527.00	105.00	126.00	61.70
	LIB106	**Ecuador**	3.53	65.93	77.39	1.50	515.00	125.00	119.33	57.20
		NL-Sc	2.24	67.97	77.06	1.64	501.33	121.33	114.67	51.23
		NL-Wi	4.82	63.89	77.71	1.35	528.67	128.67	124.00	63.17

Bibliography

Adhikari, K.N., N.W. Galwey, and M. Dracup. 2001. "The genetic control of highly restricted branching in narrow-leafed lupin (Lupinus angustifolius L.)." *Euphytica* 117 (3): 261-274. https://doi.org/10.1023/a:1026571416075. https://doi.org/10.1023/A:1026571416075.

Adhikari, Kedar N., Bevan J. Buirchell, and Mark W. Sweetingham. 2012. "Length of vernalization period affects flowering time in three lupin species." *Plant Breeding* 131 (5): 631-636. https://doi.org/10.1111/j.1439-0523.2012.01996.x.

Adomas, Barbara, Renata Galek, Gas-Smereka M, Helios W, Hurej M, Kotecki A, M. Kozak, Malarz W, Adam Okorski, Agnieszka Piotrowicz-Cieślak, Agnieszka Pszczółkowska, Ewa Sawicka-Sienkiewicz, and Jacek Twardowski. 2015. Adaptation of the Andean lupin (Lupinus mutabilis Sweet) to natural conditions of south-western Poland. 195 ed., edited by A. Kotecki: *Uniwersytet Przyrodniczy we Wrocławiu*.

Ainouche, Abdel-Kader, and Randall J. Bayer. 1999. "Phylogenetic relationships in Lupinus (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA." *American Journal of Botany* 86 (4): 590-607. https://doi.org/10.2307/2656820.

Al-Kaisey, Mahdi Thumad, and Kenneth C. B. Wilkie. 1992. "The polysaccharides of agricultural lupin seeds." *Carbohydrate Research* 227: 147-161. https://doi.org/https://doi.org/10.1016/0008-6215(92)85067-A.

Ando, Eigo, Masato Ohnishi, Yin Wang, Tomonao Matsushita, Aiko Watanabe, Yuki Hayashi, Miho Fujii, Jian Feng Ma, Shin-ichiro Inoue, and Toshinori Kinoshita. 2013. "TWIN SISTER OF FT, GIGANTEA, and CONSTANS have a positive but indirect effect on blue light-induced stomatal opening in Arabidopsis." *Plant Physiology* 162 (3): 1529-1538.

Annicchiarico, Paolo, Nathalie Harzic, and Antonio Melchiorre Carroni. 2010. "Adaptation, diversity, and exploitation of global white lupin (Lupinus albus L.) landrace genetic resources." *Field Crops Research* 119 (1): 114-124.

Annicchiarico, Paolo, Nelson Nazzicari, Tommaso Notario, Cristina Monterrubio Martin, Massimo Romani, Barbara Ferrari, and Luciano Pecetti. 2021. "Pea breeding for intercropping with cereals: variation for competitive ability and associated traits, and assessment of phenotypic and genomic selection strategies." *Frontiers in plant science* 12.

Arnold, Sarah EJ, M Eduardo Peralta Idrovo, Luis J Lomas Arias, Steven R Belmain, and Philip C Stevenson. 2014. "Herbivore defence compounds occur in pollen and reduce bumblebee colony fitness." *Journal of chemical ecology* 40 (8): 878-881.

Arnoldi, Anna, and Silvia Greco. 2011. "Nutritional and nutraceutical characteristics of lupin protein." *Nutrafoods* 10 (4): 23-29.

Atchison, Guy W, Bruno Nevado, Ruth J Eastwood, Natalia Contreras-Ortiz, Carlos Reynel, Santiago Madriñán, Dmitry A Filatov, and Colin E Hughes. 2016. "Lost crops of the Incas: Origins of domestication of the Andean pulse crop tarwi, Lupinus mutabilis." *American journal of botany* 103 (9): 1592-1606.

Atnaf, Mulugeta, Kassahun Tesfaye, and Kifle Dagne. 2015. "The Importance of Legumes in the Ethiopian Farming System and Overall Economy: An Overview." *American Journal of Experimental Agriculture* 7: 347-358. https://doi.org/10.9734/AJEA/2015/11253.

Avila, Carmen M, Salvador Nadal, M Teresa Moreno, and Ana M Torres. 2006. "Development of a simple PCR-based marker for the determination of growth habit in Vicia faba L. using a candidate gene approach." *Molecular Breeding* 17 (3): 185-190.

Ávila, Carmen M, MD Ruiz-Rodríguez, Serafin Cruz-Izquierdo, Sergio G Atienza, José I Cubero, and Ana M Torres. 2017. "Identification of plant architecture and yield-related QTL in Vicia faba L." *Molecular Breeding* 37 (7): 1-13.

Badmi, Raghuram, Raja S Payyavula, Garima Bali, Hao-Bo Guo, Sara S Jawdy, Lee E Gunter, Xiaohan Yang, Kimberly A Winkeler, Cassandra Collins, and William H Rottmann. 2018. "A new calmodulin-binding protein expresses in the context of secondary cell wall biosynthesis and impacts biomass properties in Populus." *Frontiers in plant science* 9: 1669.

Barnes, RJ, Mewa Singh Dhanoa, and Susan J Lister. 1989. "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra." *Applied spectroscopy* 43 (5): 772-777.

Bart, Rebecca S, Mawsheng Chern, Miguel E Vega-Sánchez, Patrick Canlas, and Pamela C Ronald. 2010. "Rice Snl6, a cinnamoyl-CoA reductase-like gene family member, is required for NH1-mediated immunity to Xanthomonas oryzae pv. oryzae." *PLoS genetics* 6 (9): e1001123.

Bastos Lima, Mairon G. 2018. "Toward multipurpose agriculture: food, fuels, flex crops, and prospects for a bioeconomy." *Global Environmental Politics* 18 (2): 143-150.

Belski, Regina, Trevor A Mori, Ian B Puddey, Sofia Sipsas, Richard J Woodman, Timothy R Ackland, Lawrence J Beilin, Emma R Dove, NB Carlyon, and V Jayaseena. 2011. "Effects of lupin-enriched foods on body composition and cardiovascular disease risk factors: a 12-month randomized controlled weight loss trial." *International journal of obesity* 35 (6): 810.

Benjamin, Y, H Cheng, and JF Görgens. 2013. "Evaluation of bagasse from different varieties of sugarcane by dilute acid pretreatment and enzymatic hydrolysis." *Industrial Crops and Products* 51: 7-18.

Benton, Tim G, Carling Bieg, Helen Harwatt, Roshan Pudasaini, and Laura Wellesley. 2021. "Food system impacts on biodiversity loss." *Three levers for food system transformation in support of nature. Chatham House, London.*

Berger, Jens D., Jon C. Clements, Matthew N. Nelson, Lars G. Kamphuis, Karam B. Singh, and Bevan Buirchell. 2013. "The essential role of genetic resources in narrow-leafed lupin improvement." *Crop and Pasture Science* 64 (4): 361-373, 13. https://doi.org/10.1071/CP13092.

Bermúdez-Torres, Kalina, Jorge Martínez Herrera, Rodolfo Figueroa Brito, Michael Wink, and Luc Legal. 2009. "Activity of quinolizidine alkaloids from three Mexican Lupinus against the lepidopteran crop pest Spodoptera frugiperda." *BioControl* 54 (3): 459-466.

Bertan, Ivandro, Fernando IF de Carvalho, and Antonio Costa de Oliveira. 2007. "Parental selection strategies in plant breeding programs." *Journal of crop science and biotechnology* 10 (4): 211-222.

Blanc, Guillaume, and Kenneth H Wolfe. 2004. "Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution." *The Plant Cell* 16 (7): 1679-1691.

Blanco-Galdos, O. 1982. "Genetic variability of tarwi (Lupinus mutabilis Sweet)." 25) 33-49. https://eurekamag.com/research/015/930/015930644.php.

Blanco Galdos, O. 1982. "Genetic variability of tarwi (Lupinus mutabilis Sweet)." *Schriftenreihe der Gesellschaft fur Technische Zusammenarbeit*.

Boerema, Annelies, Alain Peeters, Sanne Swolfs, Floor Vandevenne, Sander Jacobs, Jan Staes, and Patrick Meire. 2016. "Soybean trade: balancing environmental and socio-economic impacts of an intercontinental market." *PloS one* 11 (5): e0155222.

Boschin, Giovanna, and Donatella Resta. 2013. "Alkaloids Derived from Lysine: Quinolizidine (a Focus on Lupin Alkaloids)." In *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*, edited by Kishan Gopal Ramawat and Jean-Michel Mérillon, 381-403. Berlin, Heidelberg: Springer Berlin Heidelberg.

Brandon, Andrew G, and Henrik V Scheller. 2020. "Engineering of bioenergy crops: dominant genetic approaches to improve polysaccharide properties and composition in biomass." *Frontiers in plant science* 11: 282.

Brücher, H. 1989. "Lupinus mutabilis Sweet." *Useful Plants of Neotropical Origin and Their Wild Relatives, Springer-Verlag* 80.

Brummer, E Charles, Wesley T Barber, Sarah M Collier, Thomas S Cox, Randy Johnson, Seth C Murray, Richard T Olsen, Richard C Pratt, and Ann Marie Thro. 2011. "Plant breeding for harmony between agriculture and the environment." *Frontiers in Ecology and the Environment* 9 (10): 561-568.

Bunsupa, Somnuk, Kae Katayama, Emi Ikeura, Akira Oikawa, Kiminori Toyooka, Kazuki Saito, and Mami Yamazaki. 2012. "Lysine Decarboxylase Catalyzes the First Step of Quinolizidine Alkaloid Biosynthesis and Coevolved with Alkaloid Production in Leguminosae." *The Plant Cell* 24 (3): 1202. https://doi.org/10.1105/tpc.112.095885.

Caligari, P. D. S., P. Römer, M. A. Rahim, C. Huyghe, J. Neves-Martins, and E. J. Sawicka-Sienkiewicz. 2000. "The Potential of Lupinus mutabilis as a crop." In *Linking Research and Marketing Opportunities for Pulses in the 21st Century: Proceedings of the Third International Food Legumes Research Conference*, edited by R. Knight, 569-573. Dordrecht: Springer Netherlands.

Carreira, Alexandra, João Boavida Ferreira, Iliana Pereira, João Ferreira, Paulo Filipe, Ricardo Boavida Ferreira, and Sara Monteiro. 2018. "Blad-containing oligomer: a novel fungicide used in crop protection as an alternative treatment for tinea pedis and tinea versicolor." *Journal of medical microbiology* 67 (2): 198-207.

Carvajal-Larenas, F. E., A. R. Linnemann, M. J. R. Nout, M. Koziol, and M. A. J. S. van Boekel. 2016. "Lupinus mutabilis: Composition, Uses, Toxicology, and Debittering." *Critical Reviews in Food Science and Nutrition* 56 (9): 1454-1487. https://doi.org/10.1080/10408398.2013.772089.

Carvajal-Larenas, F. E. 2013. *Managing technological aspects of Lupinus mutabilis from a food sovereignty perspective in Ecuador*. PhD thesis. Wageningen (NL).

Carvalho, Isabel Saraiva de, M Chaves, and C Pinto Ricardo. 2005. "Influence of water stress on the chemical composition of seeds of two lupins (Lupinus albus and Lupinus mutabilis)." *Journal of agronomy and crop science* 191 (2): 95-98.

Chalampuente-Flores, D, CT Bastidas, and M Sørensen. 2021. "The Andean lupine-'El Chocho'or 'Tarwi'(Lupinus mutabilis Sweet). Biodivers." *Online J* 1.

Chen, Yun, Changhua Liu, Peter R Chang, Xiaodong Cao, and Debbie P Anderson. 2009. "Bionanocomposites based on pea starch and cellulose nanowhiskers hydrolyzed from pea hull fibre: effect of hydrolysis time." *Carbohydrate Polymers* 76 (4): 607-615.

Chirinos-Arias, Michelle, Jorge E. Jiménez, and Lizbeth S. Vilca-Machaca. 2015. "Analysis of Genetic Variability among thirty accessions of Andean Lupin (Lupinus mutabilis Sweet) using ISSR molecular markers." *SCIENTIA AGROPECUARIA* 6: 17-30. https://doi.org/10.17268/sci.agropecu.2015.01.02.

Christiansen, JL, and B Jørnsgård. 2002. "Influence of day length and temperature on number of main stem leaves and time to flowering in lupin." *Annals of Applied Biology* 140 (1): 29-35.

Clark, R., and S. Johnson. 2002. "Sensory Acceptability of Foods with Added Lupin (Lupinus angustifolius) Kernel Fiber Using Pre-set Criteria." *Journal of Food Science* 67 (1): 356-362. https://doi.org/10.1111/j.1365-2621.2002.tb11410.x.

Clements, JC, MS Sweetingham, L Smith, Gordon Francis, G Thomas, and S Sipsas. 2008. "Crop improvement in Lupinus mutabilis for Australian agriculture-progress and prospects." Lupins for health and wealth. Proceedings of the 12th International Lupin Conference, Fremantle, Western Australia, 14-18 September 2008.

Collins, F. I., and Robert W. Howell. 1957. "Variability of linolenic and linoleic acids in soybean oil." *Journal of the American Oil Chemists Society* 34 (10): 491-493. https://doi.org/10.1007/bf02638073.

Cowling, WA, BJ Buirchell, and ME Tapia. 1998. "Lupin. Lupinus spp. Promoting the conservation and use of underutilized and neglected crops. 23." *Institute of Plant Genetics and Crop Plant Resources: Gatersleben/International Plant Genetic Resources Institute: Rome, Italy*.

Cremer, HD. 1983. "Current aspects on legumes as a food constituent in Latin America with special emphasis on lupines: Introduction." *Plant Foods for Human Nutrition (Formerly Qualitas Plantarum)* 32 (2): 95-100.

Curti, Carolina A, Ramiro N Curti, Norberto Bonini, and Adriana N Ramón. 2018. "Changes in the fatty acid composition in bitter Lupinus species depend on the debittering process." *Food chemistry* 263: 151-154.

Czubinski, Jarosław, Anna Grygier, and Aleksander Siger. 2021. "Lupinus mutabilis seed composition and its comparison with other lupin species." *Journal of Food Composition and Analysis* 99: 103875.

Dalrymple, Dana G. 1986. *Development and spread of high-yielding rice varieties in developing countries*. Int. Rice Res. Inst.

Dargahi, Hamidreza, Patcharin Tanya, Prakit Somta, Jun Abe, and Peerasak Srinives. 2014. "Mapping quantitative trait loci for yield-related traits in soybean (Glycine max L.)." *Breeding science* 64 (4): 282-290.

Dauwe, Rebecca, Kris Morreel, Geert Goeminne, Birgit Gielen, Antje Rohde, Jos Van Beeumen, John Ralph, Alain-Michel Boudet, Joachim Kopka, and Soizic F Rochange. 2007. "Molecular phenotyping of lignin-modified tobacco reveals associated changes in cell-wall metabolism, primary metabolism, stress metabolism and photorespiration." *The Plant Journal* 52 (2): 263-285.

De Ron, Antonio, Francesca Sparvoli, José Pueyo, and Didier Bazile. 2017. "Editorial: Protein Crops: Food and Feed for the Future." *Frontiers in Plant Science* 8 (105). https://doi.org/10.3389/fpls.2017.00105.

De Vega, Jose J, Ned Peel, Sarah J Purdy, Sarah Hawkins, Lain Donnison, Sarah Dyer, and Kerrie Farrar. 2021. "Differential expression of starch and sucrose metabolic genes linked to varying biomass yield in Miscanthus hybrids." *Biotechnology for biofuels* 14 (1): 1-15.

Demurin, YA, Dragan Škorić, Ištvan Verešbaranji, and Siniša Jocić. 2000. "Inheritance of increased oleic acid content in sunflower seed oil." *Helia* 23 (32): 87-92.

Doris Chalampuente-Flores, César Tapia Bastidas, and Marten Sørensen. 2021. "The Andean Lupine-'El Chocho' or 'Tarwi' (Lupinus mutabilis Sweet)." *Biodiversity Online J.* 1 (4). https://doi.org/BOJ.000520.2021.

Dowswell, Christopher R, Ripusan L Paliwal, and Ronald P Cantrell. 2019. *Maize in the third world*. CRC press.

Doxastakis, G. 2000. "Lupin seed proteins." In *Developments in Food Science*, edited by G. Doxastakis and V. Kiosseoglou, 7-38. Elsevier.

Doyle, Jeff J, and Jan L Doyle. 1987. *A rapid DNA isolation procedure for small quantities of fresh leaf tissue.*

Dracup, Miles, Mark A Reader, and Jairo A Palta. 1998. "Variation in yield of narrow-leafed lupin caused by terminal drought." *Australian Journal of Agricultural Research* 49 (5): 799-810.

Drummond, Christopher S, Ruth J Eastwood, Silvia TS Miotto, and Colin E Hughes. 2012. "Multiple continental radiations and correlates of diversification in Lupinus (Leguminosae): testing for key innovation with incomplete taxon sampling." *Systematic biology* 61 (3): 443-460.

European Commission. 2012. Innovating for sustainable growth : a bioeconomy for Europe. edited by Directorate-General for Research and Innovation: Publications Office.

European Commission. 2018. A sustainable bioeconomy for Europe: strengthening the connection between economy, society and the environment. edited by Updated Bioeconomy Strategy. Brussels: European Commission.

Falconí, C. 2012. *Lupinus mutabilis in Ecuador with special emphasis on anthracnose resistance*.

Falter, Christian, and Christian A Voigt. 2016. "Improving biomass production and saccharification in Brachypodium distachyon through overexpression of a sucrose-phosphate synthase from sugarcane." *Journal of Plant Biochemistry and Biotechnology* 25 (3): 311-318.

Fayyaz, Hasan, and Rana Ashfaq Ahmad. 2003. "Effects of seasonal variations on oil and fatty acid profile of sunflower / influencia de las variaciones de temporada en el perfil oleaginoso y de ácidos grasos de girasol / effet des variations saisonnières sur le profil oléagineux et le profil en acides gras du tournesol." *Helia* 26 (38): 159-166. https://doi.org/10.2298/hel0338159f.

Ferreira, Helena, Elisabete Pinto, and Marta W Vasconcelos. 2021. "Legumes as a cornerstone of the transition toward more sustainable agri-food systems and diets in Europe." *Frontiers in Sustainable Food Systems* 5.

Foley, Rhonda C., Jose C. Jimenez-Lopez, Lars G. Kamphuis, James K. Hane, Su Melser, and Karam B. Singh. 2015. "Analysis of conglutin seed storage proteins across lupin species using transcriptomic, protein and comparative genomic approaches." *BMC Plant Biology* 15 (1): 106. https://doi.org/10.1186/s12870-015-0485-6.

Fornasini, M, J Castro, Elena Villacrés, L Narváez, Ma P Villamar, and Manuel Eduardo Baldeón. 2012. "Hypoglycemic effect of Lupinus mutabilis in healthy volunteers and subjects with dysglycemia." *Nutricion hospitalaria* 27 (2): 425-433.

Foucher, Fabrice, Julie Morin, Juliette Courtiade, Sandrine Cadioux, Noel Ellis, Mark J Banfield, and Catherine Rameau. 2003. "DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea." *The Plant Cell* 15 (11): 2742-2754.

Franke, Rochus, Colleen M McMichael, Knut Meyer, Amber M Shirley, Joanne C Cusumano, and Clint Chapple. 2000. "Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase." *The Plant Journal* 22 (3): 223-234.

Frick, Karen M., Lars G. Kamphuis, Kadambot H. M. Siddique, Karam B. Singh, and Rhonda C. Foley. 2017. "Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement." *Frontiers in Plant Science* 8 (87). https://doi.org/10.3389/fpls.2017.00087.

Fries, Ana María, and Mario E Tapia. 2007. *Guía de campo de los cultivos andinos*. FAO, ANPE-PERÚ.

Froyen, Erik, and Bonny Burns-Whitmore. 2020. "The effects of linoleic acid consumption on lipid risk markers for cardiovascular disease in healthy individuals: A review of human intervention trials." *Nutrients* 12 (8): 2329.

Fu, Chunxiang, Jonathan R Mielenz, Xirong Xiao, Yaxin Ge, Choo Y Hamilton, Miguel Rodriguez, Fang Chen, Marcus Foston, Arthur Ragauskas, and Joseph Bouton. 2011.

"Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass." *Proceedings of the National Academy of Sciences* 108 (9): 3803-3808.

Fujii, Satoshi, Takahisa Hayashi, and Koichi Mizuno. 2010. "Sucrose synthase is an integral component of the cellulose synthesis machinery." *Plant and cell physiology* 51 (2): 294-301.

Fund, Christin, Beate El-Chichakli, and Christian Patermann. 2018. *Bioeconomy Policy (Part III): Update Report of National Strategies Around the World.* (Berlin).

Galek, R. 2010. "Studies on the variability of some morphological and functional characters of Lupinus with particular consideration intra and interspecific hybrids." *Monografie (Poland).*

Galek, R, E Sawicka-Sienkiewicz, and D Zalewski. 2007. "Evaluation of interspecific hybrids of andean lupin and their parental forms with regard to some morphological and quantitative characters." *Fragmenta Agronomica (Poland).*

Galek, R., E. Sawicka-Sienkiewicz, D. Zalewski, S. Stawiński, and K. Spychała. 2017. "Searching for low alkaloid forms in the Andean lupin (Lupinus mutabilis) collection." *Czech Journal of Genetics and Plant Breeding* 53 (2): 55-62. https://doi.org/10.17221/71/2016-CJGPB.

Galek, Renata, Bartosz Kozak, A. Biela, Dariusz Zalewski, Ewa Sawicka-Sienkiewicz, K. Spychała, and S. Stawiński. 2016. "Seed coat thickness differentiation and genetic polymorphism for Lupinus mutabilis sweet breeding." 21: 305-312. https://doi.org/10.17557/tjfc.99967.

Garcia, Araceli, Alessandro Gandini, Jalel Labidi, Naceur Belgacem, and Julien Bras. 2016. "Industrial and crop wastes: A new source for nanocellulose biorefinery." *Industrial Crops and Products* 93: 26-38.

Gas, Marta. 2014. "Wpływ wybranych czynników agrotechnicznych na rozwój i plonowanie łubinu andyjskiego (Lupinus mutabilis Sweet)."

Ge, Shating, Xiaofei Han, Xuwen Xu, Yiming Shao, Qiankun Zhu, Yidong Liu, Juan Du, Juan Xu, and Shuqun Zhang. 2020. "WRKY15 suppresses tracheary element differentiation upstream of VND7 during xylem formation." *The Plant Cell* 32 (7) : 2307-2324.

Ghosh, Jayadri Sekhar, Shubho Chaudhuri, Nrisingha Dey, and Amita Pal. 2013. "Functional characterization of a serine-threonine protein kinase from Bambusa balcooathat implicates in cellulose overproduction and superior quality fiber formation." *BMC plant biology* 13 (1): 1-17.

Gifford, Justin M, Won Byoung Chae, Kankshita Swaminathan, Stephen P Moose, and John A Juvik. 2015. "Mapping the genome of Miscanthus sinensis for QTL associated with biomass productivity." *GCB Bioenergy* 7 (4): 797-810.

Gnatowska, M, WK Święcicki, and B Wolko. 2000. "Preliminary data on the outcrossing rate in sweet Lupinus mutabilis." Lupin, an ancient crop for the new millennium: Proceedings of the 9th International Lupin Conference, Klink/Muritz, Germany, 20-24 June, 1999.

Gocal, Greg FW, Andrew T Poole, Frank Gubler, Robyn J Watts, Cheryl Blundell, and Rod W King. 1999. "Long-day up-regulation of a GAMYB gene during Lolium temulentum inflorescence formation." *Plant Physiology* 119 (4): 1271-1278.

Gomez, Nora V., and Daniel J. Miralles. 2011. "Factors that modify early and late reproductive phases in oilseed rape (Brassica napus L.): Its impact on seed yield and oil content." *Industrial Crops and Products* 34 (2): 1277-1285. https://doi.org/https://doi.org/10.1016/j.indcrop.2010.07.013.

González-Andrés, Fernando, Pedro Casquero, Cristina San-Pedro, and Elias Hernandez. 2007. "Diversity in White Lupin (Lupinus albus L.) Landraces from Northwest Iberianplateau." *Genetic Resources and Crop Evolution* 54: 27-44. https://doi.org/10.1007/s10722-005-1407-5.

González, Ana M, Fernando J Yuste-Lisbona, Soledad Saburido, Sandra Bretones, Antonio M De Ron, Rafael Lozano, and Marta Santalla. 2016. "Major contribution of flowering time

and vegetative growth to plant production in common bean as deduced from a comparative genetic mapping." *Frontiers in plant science* 7: 1940.

Gorynowicz, B., W. Swiecicki, A. Osiecka, and Z. Kaczmarek. 2014. "Terminal inflorescence and restricted branching genes in lupins (L. albus L., L. angustifolius L., L. luteus L.) and field bean (Vicia faba L.) breeding in Poland." *Journal of Agricultural Science and Technology B* 4 (9): 712-721.

Goujon, Thomas, Valerie Ferret, Isabelle Mila, Brigitte Pollet, Katia Ruel, Vincent Burlat, Jean-Paul Joseleau, Yves Barriere, Catherine Lapierre, and Lise Jouanin. 2003. "Down-regulation of the AtCCR1 gene in Arabidopsis thaliana: effects on phenotype, lignins and cell wall degradability." *Planta* 217 (2): 218-228.

Gresta, F., M. Wink, U. Prins, M. Abberton, J. Capraro, A. Scarafoni, and G. Hill. 2017. "Lupins in European cropping systems." In *Legumes in Cropping Systems*, 88-108. CABI.

Gross, R, and E von Baer. 1981. "Lupine, eine neue Kulturpflanze in den Anden. II. Umwelt- und genotypisch bedingte Einflusse auf Olgehalt und-qualitat im Samen von Tarwi (Lupinus mutabilis Sweet)." *Zeitschrift fur Acker-und Pflanzenbau.= Journal of agronomy and crop science*.

Gross, R, E Von Baer, F Koch, R Marquard, L Trugo, and M Wink. 1988. "Chemical composition of a new variety of the Andean lupin (Lupinus mutabilis cv. Inti) with low-alkaloid content." *Journal of food composition and analysis* 1 (4): 353-361.

Guaytarilla, Pamela , and Cesar A. Falconi. 2014. "Selección pro arquitectura de la planta y rsistencia a la Antracnosis de 7 Genotipos de Chocho (Lupinus mutabilis)." IX Congreso De Ciencia Y Tecnologia 2014.

Guerriero, G., Berni, R., Muñoz-Sanchez, J. A., Apone, F., Abdel-Salam, E. M., Qahtan, A. A., & Faisal, M. 2018. "Production of plant secondary metabolites: Examples, tips and suggestions for biotechnologists". *Genes*, 9(6), 309.

Guilengue, Norberto, Sofia Alves, Pedro Talhinhas, and João Neves-Martins. 2019. "Genetic and genomic diversity in a tarwi (Lupinus mutabilis Sweet) germplasm collection and adaptability to Mediterranean climate conditions." *Agronomy* 10 (1): 21.

Guilengue, Norberto, João Neves-Martins, and Pedro Talhinhas. 2020. "Response to anthracnose in a tarwi (Lupinus mutabilis) collection is influenced by anthocyanin pigmentation." *Plants* 9 (5): 583.

Gulisano, Agata, Sofia Alves, João Neves Martins, and Luisa M. Trindade. 2019c. "Genetics and Breeding of Lupinus mutabilis: An Emerging Protein Crop." *Frontiers in Plant Science* 10 (1385). https://doi.org/10.3389/fpls.2019.01385.

Gulisano, Agata, Sofia Alves, Diego Rodriguez, Angel Murillo, Bert-Jan Van Dinter, Andres F Torres, Milton Gordillo-Romero, Maria de Lourdes Torres, João Neves-Martins, and Maria-João Paulo. 2022. "Diversity and Agronomic Performance of Lupinus mutabilis Germplasm in European and Andean Environments." *Frontiers in Plant Science* 13.

Gupta, Pushpendra Kumar, Sachin Rustgi, and Pawan L. Kulwal. 2005. "Linkage disequilibrium and association studies in higher plants: Present status and future prospects." *Plant Molecular Biology* 57: 461-485.

Hackbarth, J. 1961. "Lupinosis in the light of old and new evidence." *Journal ofthe Australian Institute ofAgrlcultura1 Science* 27 (6): 1-7.

Hamama, Anwar A, and Harbans L Bhardwaj. 2004. "Phytosterols, triterpene alcohols, and phospholipids in seed oil from white lupin." *Journal of the American Oil Chemists' Society* 81 (11): 1039-1044.

Hane, James K., Yao Ming, Lars G. Kamphuis, Matthew N. Nelson, Gagan Garg, Craig A. Atkins, Philipp E. Bayer, Armando Bravo, Scott Bringans, Steven Cannon, David Edwards, Rhonda Foley, Ling-ling Gao, Maria J. Harrison, Wei Huang, Bhavna Hurgobin,

Sean Li, Cheng-Wu Liu, Annette McGrath, Grant Morahan, Jeremy Murray, James Weller, Jianbo Jian, and Karam B. Singh. 2017. "A comprehensive draft genome sequence for lupin (Lupinus angustifolius), an emerging health food: insights into plant–microbe interactions and legume evolution." *Plant Biotechnology Journal* 15 (3): 318-330. https://doi.org/10.1111/pbi.12615.

Haq, N. 1993. "Lupins (Lupinus species)." *Pulses and vegetables*: 103-129.

Hardy, A, C Huyghe, and J Papineau. 1997. "Dry matter accumulation and partitioning, and seed yield in indeterminate Andean lupin (Lupinus mutabilis Sweet)." *Australian Journal of Agricultural Research* 48 (1): 91-102.

Hardy, A., C. Huyghe, M. A. Rahim, P. Roemer, J. M. Neves-Martins, E. Sawicka-Sienkiewicz, and P. D. S. Caligari. 1998. "Effects of genotype and environment on architecture and flowering time of indeterminate Andean lupins (Lupinus mutabilis Sweet)." *Australian Journal of Agricultural Research* 49 (8): 1241-1252. https://doi.org/https://doi.org/10.1071/A98060.

Harzic, N, Christian Huyghe, J Papineau, Claire Billot, Robert Esnault, and C Deroo. 1996. "Genotypic variation of seed yield and architectural traits in dwarf autumn-sown white lupin." *Agronomie* 16 (5): 309-319.

He, Zuhua, Qun Zhu, Tsegaye Dabi, Debao Li, Detlef Weigel, and Chris Lamb. 2000. "Transformation of rice with the Arabidopsis floral regulator LEAFY causes early heading." *Transgenic research* 9 (3): 223-227.

Ho-Yue-Kuang, Séverine, Camille Alvarado, Sébastien Antelme, Brigitte Bouchet, Laurent Cézard, Philippe Le Bris, Frédéric Legée, Alessandra Maia-Grondard, Arata Yoshinaga, and Luc Saulnier. 2016. "Mutation in Brachypodium caffeic acid O-methyltransferase 6 alters stem and grain lignins and improves straw saccharification without deteriorating grain quality." *Journal of experimental botany* 67 (1): 227-237.

Hodgson-Kratky, K, V Perlo, A Furtado, H Choudhary, JM Gladden, BA Simmons, F Botha, and RJ Henry. 2021. "Association of gene expression with syringyl to guaiacyl ratio in sugarcane lignin." *Plant Molecular Biology* 106 (1): 173-192.

Hodgson, Edward M, Susan J Lister, Anthony V Bridgwater, John Clifton-Brown, and Iain S Donnison. 2010. "Genotypic and environmentally derived variation in the cell wall composition of Miscanthus in relation to its use as a biomass feedstock." *Biomass and Bioenergy* 34 (5): 652-660.

Hudson, Bertram J. F., John G. Fleetwood, and Joseph I. Lewis. 1983. "Oil Content, Fatty Acids And Unsaponifiable Lipids Of Lupin Seed." *Journal of Plant Foods* 5 (1): 15-21. https://doi.org/10.1080/0142968X.1983.11904271.

Hufnagel, Barbara, André Marques, William MARANDE, Erika Sallet, Alexandre Sorriano, Sandrine Arribat, Helene Berges, Jerome Gouzy, and Benjamin PERET. 2018. "Genome sequence of white lupin, a model to study root developmental adaptations." 12th Congress of the International Plant Molecular Biology, Montpellier, France, 2018-08-05.

Hughes, Colin, and Ruth Eastwood. 2006. "Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes." *Proceedings of the National Academy of Sciences* 103 (27): 10334-10339. https://doi.org/10.1073/pnas.0601928103.

Hughes, Stephen R, William R Gibbons, Bryan R Moser, and Joseph O Rich. 2013. "Sustainable multipurpose biorefineries for third-generation biofuels and value-added co-products." *Biofuels-economy, environment and sustainability* 2013: 245-262.

Huyghe, Christian. 1998. "Genetics and genetic modifications of plant architecture in grain legumes: a review." *Agronomie* 18 (5-6): 383-411.

Iqbal, Muhammad Javed, Sujan Mamidi, Rubina Ahsan, Shahryar F. Kianian, Clarice J. Coyne, Anwar A. Hamama, Satya S. Narina, and Harbans L. Bhardwaj. 2012. "Population

structure and linkage disequilibrium in Lupinus albus L. germplasm and its implication for association mapping." *Theoretical and applied genetics* 125 (3): 517-530.

Iqbal, Yasir, Iris Lewandowski, Axel Weinreich, Bernd Wippel, Barbara Pforte, Olha Hadai, Oleksandra Tryboi, Matthias Spöttle, and Daan Peters. 2016. *Maximising the yield of biomass from residues of agricultural crops and biomass from forestry*. Ecofys.

Jacobsen, Sven-Erik, and Angel Mujica. 2006. "El Tarwi (Lupinus mutabilis Sweet.) y sus parientes silvestres." *Botánica Económica de los Andes Centrales* 28.

Jha, Uday C, Harsh Nayyar, Swarup K Parida, Rupesh Deshmukh, Eric JB von Wettberg, and Kadambot HM Siddique. 2022. "Ensuring Global Food Security by Improving Protein Content in Major Grain Legumes Using Breeding and 'Omics' Tools." *International Journal of Molecular Sciences* 23 (14): 7710.

Jing, B., Ishikawa, T., Soltis, N., Inada, N., Liang, Y., Murawska, G., Andeberhan, F., Pidatala, R., Yu, X., Baidoo, E. and Kawai-Yamada, M., 2018. GONST2 transports GDP-Mannose for sphingolipid glycosylation in the Golgi apparatus of Arabidopsis. *BioRxiv*, p.346775.

Jones, Louise, A Roland Ennos, and Simon R Turner. 2001. "Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis." *The Plant Journal* 26 (2): 205-216.

Jung, Je Hyeong, Wilfred Vermerris, Maria Gallo, Jeffrey R Fedenko, John E Erickson, and Fredy Altpeter. 2013. "RNA interference suppression of lignin biosynthesis increases fermentable sugar yields for biofuel production from field-grown sugarcane." *Plant biotechnology journal* 11 (6): 709-716.

Kaler, Avjinder S, and Larry C Purcell. 2019. "Estimation of a significance threshold for genome-wide association studies." *BMC genomics* 20 (1): 1-8.

Kambhampati, Shrikaar, Jose A Aznar-Moreno, Cooper Hostetler, Tara Caso, Sally R Bailey, Allen H Hubbard, Timothy P Durrett, and Doug K Allen. 2019. "On the inverse correlation of protein and oil: Examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants." *Metabolites* 10 (1): 18.

Kamm, B, M Kamm, M Schmidt, I Starke, and E Kleinpeter. 2006. "Chemical and biochemical generation of carbohydrates from lignocellulose-feedstock (Lupinus nootkatensis)—quantification of glucose." *Chemosphere* 62 (1): 97-105.

Kamm, Birgit, and Michael Kamm. 2007. "International biorefinery systems." *Pure and Applied Chemistry* 79 (11): 1983-1997.

Kang, Hyun Min, Chun Ye, and Eleazar Eskin. 2008. "Accurate discovery of expression quantitative trait loci under confounding from spurious and genuine regulatory hotspots." *Genetics* 180 (4): 1909-1925.

Kang, Hyun Min, Noah A Zaitlen, Claire M Wade, Andrew Kirby, David Heckerman, Mark J Daly, and Eleazar Eskin. 2008. "Efficient control of population structure in model organism association mapping." *Genetics* 178 (3): 1709-1723.

Käss, Ernst, and Michael Wink. 1997. "Molecular phylogeny and phylogeography ofLupinus (Leguminosae) inferred from nucleotide sequences of therbcL gene and ITS 1 + 2 regions of rDNA." *Plant Systematics and Evolution* 208 (3): 139-167.https://doi.org/10.1007/bf00985439.

Keller, J., M. Rousseau-Gueutin, G. E. Martin, J. Morice, J. Boutte, E. Coissac, M. Ourari, M. Aïnouche, A. Salmon, F. Cabello-Hurtado, and A. Aïnouche. 2017. "The evolutionary fate of the chloroplast and nuclear rps16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from Lupinus." *DNA research : an international journal for rapid publication of reports on genes and genomes* 24 (4): 343-358. https://doi.org/10.1093/dnares/dsx006

Kerstens M. H., Schranz M. E., Bouwmeester K. 2020. Phylogenomic analysis of the APETALA2 transcription factor subfamily across angiosperms reveals both deep conservation and lineage-specific patterns. *Plant Journal.* 103 1516–1524. 10.1111/tpj.14843

Klein, Anthony, Hervé Houtin, Céline Rond-Coissieux, Myriam Naudet-Huart, Michael Touratier, Pascal Marget, and Judith Burstin. 2020. "Meta-analysis of QTL reveals the genetic control of yield-related traits and seed protein content in pea." *Scientific reports* 10 (1): 1-11.

Korte, Arthur, and Ashley Farlow. 2013. "The advantages and limitations of trait analysis with GWAS: a review." *Plant methods* 9 (1): 1-9.

Kroc, Magdalena, Grzegorz Koczyk, Wojciech Święcicki, Andrzej Kilian, and Matthew Nelson. 2014. "New evidence of ancestral polyploidy in the Genistoid legume Lupinus angustifolius L. (narrow-leafed lupin)." *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* 127. https://doi.org/10.1007/s00122-014-2294-y.

Krylova, E.A., E.K. Khlestkina, M.O. Burlyaeva, and M.A. Vishnyakova. 2020. "Determinate growth habit of grain legumes: role in domestication and selection, genetic control." *Ecological genetics* 18 (1): 43-58. https://doi.org/10.17816/ecogen16141.

Książkiewicz, Michał, Nelson Nazzicari, Hua'an Yang, Matthew N Nelson, Daniel Renshaw, Sandra Rychel, Barbara Ferrari, Maria Carelli, Magdalena Tomaszewska, and Stanisław Stawiński. 2017. "A high-density consensus linkage map of white lupin highlights synteny with narrow-leafed lupin and provides markers tagging key agronomic traits." *Scientific Reports* 7 (1): 1-15.

Kumar, Parveen, Diane M Barrett, Michael J Delwiche, and Pieter Stroeve. 2009. "Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production." *Industrial & engineering chemistry research* 48 (8): 3713-3729.

Kumar, S, and S Gugita. 2004. "How wide the genetic base of pulse crops is. Pulses in new perspective. Eds.: Ali M, Singh BB, Kumar S, Dhar V." *Proc. Nat. Simp. Crop Diversification Natur. Resour. Manag. Kanpur, India*: 211-217.

Kwak, Myounghai, Dianne Velasco, and Paul Gepts. 2008. "Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (Phaseolus vulgaris)." *Journal of Heredity* 99 (3): 283-291.

Lambers, Hans, Jon C Clements, and Matthew N Nelson. 2013. "How a phosphorus-acquisition strategy based on carboxylate exudation powers the success and agronomic potential of lupines (Lupinus, Fabaceae)." *American Journal of Botany* 100 (2): 263-288.

Lazaridi, Efstathia, George K Papadopoulos, and Penelope J Bebeli. 2020. "Andean lupin phenology and agronomic performance under different planting dates in a mediterranean climate." *Agronomy* 10 (12).

Li, Kun, Hongwu Wang, Xiaojiao Hu, Zhifang Liu, Yujin Wu, and Changling Huang. 2016. "Genome-wide association study reveals the genetic basis of stalk cell wall components in maize." *PLoS One* 11 (8): e0158906.

Li, Mi, Yunqiao Pu, and Arthur J Ragauskas. 2016. "Current understanding of the correlation of lignin structure with biomass recalcitrance." *Frontiers in chemistry* 4: 45.

Li, Xia, Li Chen, Meiyan Hong, Yan Zhang, Feng Zu, Jing Wen, Bin Yi, Chaozhi Ma, Jinxiong Shen, and Jinxing Tu. 2012. "A large insertion in bHLH transcription factor BrTT8 resulting in yellow seed coat in Brassica rapa." *PLoS One* 7 (9): e44145.

Li, Zhaobo, Qun Cheng, Zhuoran Gan, Zhihong Hou, Yuhang Zhang, Yongli Li, Haiyang Li, Haiyang Nan, Cen Yang, and Linnan Chen. 2021. "Multiplex CRISPR/Cas9-mediated knockout of soybean LNK2 advances flowering time." *The Crop Journal* 9 (4): 767-776.

Li, Zicong, Yang Ou, Zhicheng Zhang, Jianming Li, and Yuehui He. 2018. "Brassinosteroid signaling recruits histone 3 lysine-27 demethylation activity to FLOWERING LOCUS C chromatin to inhibit the floral transition in Arabidopsis." *Molecular plant* 11 (9): 1135-1146.

LIBBIO. 2017. "Lupinus mutabilis for Increased Biomass from marginal lands and value for BIOrefineries." Accessed 22-07-2022. https://libbio.net/.

Lima-Cabello, Elena, Victor Alche, Rhonda C. Foley, Sofianos Andrikopoulos, Grant Morahan, Karam B. Singh, Juan D. Alche, and Jose C. Jimenez-Lopez. 2017. "Narrow-leafed lupin (Lupinus angustifolius L.) β-conglutin proteins modulate the insulin signaling pathway as potential type 2 diabetes treatment and inflammatory-related disease amelioration." *Molecular Nutrition & Food Research* 61 (5): 1600819. https://doi.org/10.1002/mnfr.201600819.

Lionetti, Vincenzo, Fedra Francocci, Simone Ferrari, Chiara Volpi, Daniela Bellincampi, Roberta Galletti, Renato D'Ovidio, Giulia De Lorenzo, and Felice Cervone. 2010. "Engineering the cell wall by reducing de-methyl-esterified homogalacturonan improves saccharification of plant tissues for bioconversion." *Proceedings of the National Academy of Sciences* 107 (2): 616-621.

Lizarazo, Clara, Frederick Stoddard, Pirjo Mäkelä, and Arja Santanen. 2010. "Genetic variability in the physiological responses of Andean lupin to drought stress." *Suomen Maataloustieteellisen Seuran Tiedote*: 1-5. https://doi.org/10.33354/smst.76862.

Lo´ pez-Bellido, Luis, Mariano Fuentes, and Juan E Castillo. 2000. "Growth and yield of white lupin under Mediterranean conditions: Effect of plant density." *Agronomy Journal* 92 (2): 200-205.

López-Malvar, Ana, A Butron, RA Malvar, SJ McQueen-Mason, L Faas, LD Gómez, P Revilla, David Jose Figueroa-Garrido, and R Santiago. 2021. "Association mapping for maize stover yield and saccharification efficiency using a multiparent advanced generation intercross (MAGIC) population." *Scientific reports* 11 (1): 1-9.

Lucas, M. Mercedes, Frederick Stoddard, Paolo Annicchiarico, Juana Frias, Cristina Martinez-Villaluenga, Daniela Sussmann, Marcello Duranti, Alice Seger, Peter Zander, and José Pueyo. 2015. "The future of lupin as a protein crop in Europe." *Frontiers in Plant Science* 6 (705). https://doi.org/10.3389/fpls.2015.00705.

Magni, Chiara, Fabio Sessa, Elena Accardo, Marco Vanoni, Paolo Morazzoni, Alessio Scarafoni, and Marcello Duranti. 2004. "Conglutin γ, a lupin seed protein, binds insulin in vitro and reduces plasma glucose levels of hyperglycemic rats." *The Journal of Nutritional Biochemistry* 15 (11): 646-650. https://doi.org/https://doi.org/10.1016/j.jnutbio.2004.06.009.

Malik, Pernille L, Luc Janss, Linda K Nielsen, Finn Borum, Henning Jørgensen, Birger Eriksen, Jan K Schjoerring, and Søren K Rasmussen. 2019. "Breeding for dual-purpose wheat varieties using marker–trait associations for biomass yield and quality traits." *Theoretical and Applied Genetics* 132 (12): 3375-3398.

Marriott, Poppy E, Leonardo D Gómez, and Simon J McQueen-Mason. 2016. "Unlocking the potential of lignocellulosic biomass through plant science." *New phytologist* 209 (4): 1366-1381.

Martins, João Manuel Neves, Pedro Talhinhas, and Raul Bruno de Sousa. 2016. "Yield and seed chemical composition of Lupinus mutabilis in Portugal." *Revista de Ciências Agrárias* 39: 518-525. http://www.scielo.mec.pt/scielo.php?script=sci_arttext&pid=S0871-018X2016000400005&nrm=iso.

Masarin, Fernando, Daniela B Gurpilhares, David CF Baffa, Márcio HP Barbosa, Walter Carvalho, André Ferraz, and Adriane MF Milagres. 2011. "Chemical composition and

enzymatic digestibility of sugarcane clones selected for varied lignin content." *Biotechnology for biofuels* 4 (1): 1-10.

Masefield, GB. 1976. "Further trials of pearl lupins in England." *Experimental Agriculture* 12 (2): 97-102.

Méchin, Valérie, Odile Argillier, Yannick Hébert, Emmanuelle Guingo, Laurence Moreau, Alain Charcosset, and Yves Barriere. 2001. "Genetic analysis and QTL mapping of cell wall digestibility and lignification in silage maize." *Crop Science* 41 (3): 690-697.

Mercado, Geovana, Jhaquelin Davalos, Ipdrs, Hivos, and Cipca. 2018. *Los caminos del tarwi y la integración andina: Bolivia, Perú y Ecuador*.

Mikel, Mark A, Brian W Diers, Randall L Nelson, and Hebron H Smith. 2010. "Genetic diversity and agronomic improvement of North American soybean germplasm." *Crop science* 50 (4): 1219-1229.

Mikić, Aleksandar, Branko Ćupina, Vojislav Mihailović, Đorđe Krstić, Svetlana Antanasović, Lana Zorić, Vuk Đorđević, Vesna Perić, and Mirjana Srebrić. 2013. "Intercropping white (Lupinus albus) and Andean (Lupinus mutabilis) lupins with other annual cool season legumes for forage production." *South African Journal of Botany* 89: 296-300.

Monyem, Abdul, and Jon H. Van Gerpen. 2001. "The effect of biodiesel oxidation on engine performance and emissions." *Biomass and Bioenergy* 20 (4): 317-325. https://doi.org/https://doi.org/10.1016/S0961-9534(00)00095-7.

Mosse, Jacques. 1990. "Nitrogen-to-protein conversion factor for ten cereals and six legumes or oilseeds. A reappraisal of its definition and determination. Variation according to species and to seed protein content." *Journal of Agricultural and Food Chemistry* 38 (1): 18-24. https://doi.org/10.1021/jf00091a004.

Mousavi-Derazmahalleh, Mahsa, Philipp E. Bayer, Bruno Nevado, Bhavna Hurgobin, Dmitry Filatov, Andrzej Kilian, Lars G. Kamphuis, Karam B. Singh, Jens D. Berger, James K. Hane, David Edwards, William Erskine, and Matthew N. Nelson. 2018. "Exploring the genetic and adaptive diversity of a pan-Mediterranean crop wild relative: narrow-leafed lupin." *Theoretical and Applied Genetics* 131 (4): 887-901. https://doi.org/10.1007/s00122-017-3045-7.

Mousavi-Derazmahalleh, Mahsa, Bruno Nevado, Philipp E Bayer, Dmitry A Filatov, James K Hane, David Edwards, William Erskine, and Matthew N Nelson. 2018. "The western Mediterranean region provided the founder population of domesticated narrow-leafed lupin." *Theoretical and Applied Genetics* 131 (12): 2543-2554.

Muzquiz, Mercedes, Alejandro Varela, Carmen Burbano, Carmen Cuadrado, Eva Guillamón, and Mercedes M Pedrosa. 2012. "Bioactive compounds in legumes: pronutritive and antinutritive actions. Implications for nutrition and health." *Phytochemistry reviews* 11 (2): 227-244.

Naganowska, Barbara, Bogdan Wolko, Elwira Sliwinska, and Zygmunt Kaczmarek. 2003. "Nuclear DNA Content Variation and Species Relationships in the Genus Lupinus (Fabaceae)." *Annals of botany* 92: 349-55. https://doi.org/10.1093/aob/mcg145.

Naghshi, Sina, Dagfinn Aune, Joseph Beyene, Sara Mobarak, Masoomeh Asadi, and Omid Sadeghi. 2021. "Dietary intake and biomarkers of alpha linolenic acid and risk of all cause, cardiovascular, and cancer mortality: systematic review and dose-response meta-analysis of cohort studies." *bmj* 375.

Nakamura, M. and Hashimoto, T., 2020. Mechanistic insights into plant chiral growth. *Symmetry*, *12*(12), p.2056.

Nelson, Matthew N, Michał Książkiewicz, Sandra Rychel, Naghmeh Besharat, Candy M Taylor, Katarzyna Wyrwa, Ricarda Jost, William Erskine, Wallace A Cowling, and Jens D Berger. 2017. "The loss of vernalization requirement in narrow-leafed lupin is associated with a

deletion in the promoter and de-repressed expression of a Flowering Locus T (FT) homologue." *New Phytologist* 213 (1): 220-232.

Nguyen, Duong T, Leonardo D Gomez, Andrea Harper, Claire Halpin, Robbie Waugh, Rachael Simister, Caragh Whitehead, Helena Oakey, Huong T Nguyen, and Tuat V Nguyen. 2020. "Association mapping identifies quantitative trait loci (QTL) for digestibility in rice straw." *Biotechnology for biofuels* 13 (1): 1-16.

Nordborg, Magnus, Justin O Borevitz, Joy Bergelson, Charles C Berry, Joanne Chory, Jenny Hagenblad, Martin Kreitman, Julin N Maloof, Tina Noyes, and Peter J Oefner. 2002. "The extent of linkage disequilibrium in Arabidopsis thaliana." *Nature genetics* 30 (2): 190-193.

Ochoa-Zavala, Maried, Pilar Suárez-Montes, Mariana Chávez-Pesqueira, Diana E López-Cobos, Carmen Julia Figueredo, Alegría David, Hilares Raquel, Yandira Lazo, Stefanía Sibille, and Rosa Villanueva. 2016. "Diferencias en morfología y visitantes florales entre Lupinus mutabilis y Lupinus aff. ballianus en la microcuenca de Warmiragra, Perú." *Chiapas* 353: 356.

Odegard, Ingrid, Harry Croezen, and Geert Bergsma. 2012. "Cascading of Biomass. 13 Solutions for a Sustainable Bio-based Economy. Making Better Choices for Use of Biomass Residues, By-products and Wastes."

Olczak, Teresa, Michal Rurek, Hanna Janska, H. Augustyniak, and Ewa Sawicka-Sienkiewicz. 2001. "Screening of cytoplasmic DNA diversity between and within Lupinus mutabilis Sweet and Lupinus albus sensu lato by restriction fragment length polymorphism (RFLP)." *Journal of applied genetics* 42: 127-37.

Osorio, Claudia E., Joshua A. Udall, Haroldo Salvo-Garrido, and Iván J. Maureira-Butler. 2018. "Development and characterization of InDel markers for Lupinus luteus L. (Fabaceae) and cross-species amplification in other Lupin species." *Electronic Journal of Biotechnology* 31: 44-47. https://doi.org/https://doi.org/10.1016/j.ejbt.2017.11.002.

Pancaldi, Francesco, and Luisa M Trindade. 2020. "Marginal lands to grow novel bio-based crops: A plant breeding perspective." *Frontiers in plant science* 11: 227.

Parra-González, Lorena B., Gabriela A. Aravena-Abarzúa, Cristell S. Navarro-Navarro, Joshua Udall, Jeff Maughan, Louis M. Peterson, Haroldo E. Salvo-Garrido, and Iván J. Maureira-Butler. 2012. "Yellow lupin (Lupinus luteus L.) transcriptome sequencing: molecular marker development and comparative studies." *BMC Genomics* 13 (1): 425. https://doi.org/10.1186/1471-2164-13-425..

Pate, JS, W Williams, and P Farrington. 1985. "Lupin (Lupinus spp.)."

Pennells, Jordan, Alan Cruickshank, Céline Chaléat, Ian D Godwin, and Darren J Martin. 2021. "Sorghum as a novel biomass for the sustainable production of cellulose nanofibers." *Industrial Crops and Products* 171: 113917.

Peralta, Eduardo, Ángel Murillo, Marco Rivera, Diego Rodríguez, Luis Lomas, and Carlos Monar. 2012. "Chocho, quinua, amaranto y ataco." *Manual agrícola de granos andinos*: 4-5.

Perez, J.F., R.D. Quidiello, and M.V.T. Sainz. 1984. "Determinacion de la produccion y contenidos proteico y graso de 217 ecotipos de L. mutabilis Sweet." Proc. 3rd Int. Lupin Conf., La Rochelle, France.

Petit, Jordi, Elma MJ Salentijn, Maria-João Paulo, Christel Denneboom, Eibertus N Van Loo, and Luisa M Trindade. 2020. "Elucidating the genetic architecture of fiber quality in hemp (Cannabis sativa L.) using a genome-wide association study." *Frontiers in genetics*: 1101.

Petit, Jordi, Elma MJ Salentijn, Maria-João Paulo, Claire Thouminot, Bert Jan van Dinter, Gianmaria Magagnini, Hans-Jörg Gusovius, Kailei Tang, Stefano Amaducci, and Shaoliang Wang. 2020. "Genetic variability of morphological, flowering, and biomass quality traits in hemp (Cannabis sativa L.)." *Frontiers in plant science* 11: 102.

Podlesny, J, and A Podlesna. 2011. "Effect of rainfall amount and distribution on growth, development and yields of determinate and indeterminate cultivars of blue lupin." *Polish Journal of Agronomy* (04).

Prakash, Dhan, and P. S. Misra. 1988. "Protein content and amino acid profile of some wild leguminous seeds." *Plant Foods for Human Nutrition* 38 (1): 61-65. https://doi.org/10.1007/bf01092311.

Przybylak, Jacek, Danuta Ciesiołka, Waleria Wysocka, Pedro M. García-López, Mario Ruiz, Wojciech Wysocki, and Krzysztof Gulewicz. 2005. "Alkaloid profiles of Mexican wild lupin and an effect of alkaloid preparation from Lupinus exaltatus seeds on growth and yield of paprika (Capsicum annuum L.)." *Industrial Crops and Products* 21: 1-7. https://doi.org/10.1016/j.indcrop.2003.12.001.

Qian, W, O Sass, J Meng, M Li, M Frauen, and C Jung. 2007. "Heterotic patterns in rapeseed (Brassica napus L.): I. Crosses between spring and Chinese semi-winter lines." *Theoretical and Applied Genetics* 115 (1): 27-34.

Quijada, Pablo A, Joshua A Udall, Hieronim Polewicz, Robert D Vogelzang, and Thomas C Osborn. 2004. "Phenotypic effects of introgressing French winter germplasm into hybrid spring canola." *Crop Science* 44 (6): 1982-1989.

Radillo, J Jesús Vargas, Mario A Ruiz-López, Ramón Rodríguez Macías, Lucía Barrientos Ramírez, Pedro M García-López, and Fernando A López-Dellamary Toral. 2011. "Fermentable sugars from lupinus rotundiflorus biomass by hydrochloric acid hydrolysis." *Bioresources* 6 (1): 344-355.

Ranocha, Philippe, Nicolas Denancé, Ruben Vanholme, Amandine Freydier, Yves Martinez, Laurent Hoffmann, Lothar Köhler, Cécile Pouzet, Jean-Pierre Renou, and Björn Sundberg. 2010. "Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers." *The Plant Journal* 63 (3): 469-483.

Reader, MA, M Dracup, and EJM Kirby. 1995. "Time to flowering in narrow-leafed lupin." *Australian Journal of Agricultural Research* 46 (5): 1063-1077.

Remington, David L, Jeffry M Thornsberry, Yoshihiro Matsuoka, Larissa M Wilson, Sherry R Whitt, John Doebley, Stephen Kresovich, Major M Goodman, and Edward S Buckler. 2001. "Structure of linkage disequilibrium and phenotypic associations in the maize genome." *Proceedings of the National Academy of Sciences* 98 (20): 11479-11484.

Ren, Longhui, Wei Huang, and Steven B Cannon. 2019. "Reconstruction of ancestral genome reveals chromosome evolution history for selected legume species." *New Phytologist* 223 (4): 2090-2103.

Renaud, Erica NC, Edith T Lammerts van Bueren, James R Myers, Maria João Paulo, Fred A Van Eeuwijk, Ning Zhu, and John A Juvik. 2014. "Variation in broccoli cultivar phytochemical content under organic and conventional management systems: Implications in breeding for nutrition." *PLoS One* 9 (7): e95683.

Robbins, Matthew D, Sung-Chur Sim, Wencai Yang, Allen Van Deynze, Esther van der Knaap, Tarek Joobeur, and David M Francis. 2011. "Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato." *Journal of experimental botany* 62 (6): 1831-1845.

Rodriguez-Alvarez, Maria Xose, Martin P Boer, Fred A van Eeuwijk, and Paul HC Eilers. 2018. "Correcting for spatial heterogeneity in plant breeding experiments with P-splines." *Spatial Statistics* 23: 52-71.

Romeo, Flora, Simona Fabroni, Gabriele Ballistreri, Serena Muccilli, Spina Alfio, and Paolo Rapisarda. 2018. "Characterization and Antimicrobial Activity of Alkaloid Extracts from

Seeds of Different Genotypes of Lupinus spp." *Sustainability* 10: 788. https://doi.org/10.3390/su10030788.

Romer, P. 1995. "New attempts to select early maturing Lupinus mutabilis for middle Europe." Proceedings of the 2nd European Conference on Grain Legumes, Copenhagen, Denmark.

Römer, P. 1994. "A determinated mutant of L. mutabilis as a possible source of early maturity." Advances in Lupin research. 7th International Lupin conference, Evora.

Romer, P, and W Jahn-Deesbach. 1992. "Eight years of experiences in breeding Lupinus mutabilis under Middle European conditions." *Agrimed Research Programme-Lupinus mutabilis: its adaptation and production under European pedoclimatic conditions (Commission of the European Communities). EUR* 14102: 79-85.

Romer, P, and W John-Deesbach. 1986. "Developments in breeding of Lupinus mutabilis.[Conference paper]." 4. International Lupin Conference, Geraldton, WA (Australia), 17 Aug 1986.

Rubenschuh, U. 1997. "[On the variability of yield components and seed components of Lipinus mutabilis].[German]."

Ruel, Katia, Jimmy Berrio-Sierra, Mohammad Mir Derikvand, Brigitte Pollet, Johanne Thévenin, Catherine Lapierre, Lise Jouanin, and Jean-Paul Joseleau. 2009. "Impact of CCR1 silencing on the assembly of lignified secondary walls in Arabidopsis thaliana." *New Phytologist* 184 (1): 99-113.

Rugnone, Matias L, Ana Faigón Soverna, Sabrina E Sanchez, Ruben Gustavo Schlaen, Carlos Esteban Hernando, Danelle K Seymour, Estefanía Mancini, Ariel Chernomoretz, Detlef Weigel, and Paloma Más. 2013. "LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator." *Proceedings of the National Academy of Sciences* 110 (29): 12120-12125.

Ruiz-López, Mario Alberto, Lucia Barrientos-Ramírez, Pedro Macedonio García-López, Elia Herminia Valdés-Miramontes, Juan Francisco Zamora-Natera, Ramón Rodríguez-Macias, Eduardo Salcedo-Pérez, Jacinto Bañuelos-Pineda, and J Jesús Vargas-Radillo. 2019. "Nutritional and bioactive compounds in mexican lupin beans species: A mini-review." *Nutrients* 11 (8): 1785.

Rychel, Sandra, Michał Książkiewicz, Magdalena Tomaszewska, Wojciech Bielski, and Bogdan Wolko. 2019. "FLOWERING LOCUS T, GIGANTEA, SEPALLATA, and FRIGIDA homologs are candidate genes involved in white lupin (Lupinus albus L.) early flowering." *Molecular Breeding* 39 (3): 1-17.

Salentijn, Elma MJ, Qingying Zhang, Stefano Amaducci, Ming Yang, and Luisa M Trindade. 2015. "New developments in fiber hemp (Cannabis sativa L.) breeding." *Industrial crops and products* 68: 32-41.

Santana, Filomena C., and J. Empis. 2001. "Bacterial removal of quinolizidine alkaloids from Lupinus albus flours." *European Food Research and Technology* 212 (2): 217-224. https://doi.org/10.1007/s002170000221.

Santos, Cláudia N., Ricardo B. Ferreira, and Artur R. Teixeira. 1997. "Seed Proteins of Lupinus mutabilis." *Journal of Agricultural and Food Chemistry* 45 (10): 3821-3825. https://doi.org/10.1021/jf970075v.

Sawicka-Sienkiewicz, E., R. Galek, H. Kalinska, D. Zalewski, and S. Stawiński. 2006. "Proportion of pod wall and seed coat in mutants of cv. Emir and in a collection of Lupinus angustifolius." 2006.

Sawicka-Sienkiewicz, E. J., and W. Kadlubiec. 2001. "Aktualny stan badan nad lubinem andyjskim [Lupinus mutabilis Sweet]." *Zeszyty Naukowe Akademii Rolniczej we Wrocławiu. Rolnictwo* 82: 115-129.

Sawicka, EJ. 1993. "The induced mutations in Andean lupine (Lupinus mutabilis Sweet)." *Prace Ogrodu Bot., PAN, s. Monografie i Rozprawy* 3: 1-112.

Schnitzer, Morris, Carlos M Monreal, and Erin E Powell. 2014. "Wheat straw biomass: A resource for high-value chemicals." *Journal of Environmental Science and Health, Part B* 49 (1): 51-67.

Schoeneberger, H., R. Gross, H. D. Cremer, and I. Elmadfa. 1982. "Composition and Protein Quality of Lupinus Mutabilis." *The Journal of Nutrition* 112 (1): 70-76. https://doi.org/10.1093/jn/112.1.70.

Schoenmakere, M de, Ybele Hoogeveen, Jeroen Gillabel, and Saskia Manshoven. 2018. "The circular economy and the bioeconomy: partners in sustainability." *EEA Report* (8/2018).

Selle, Maria Lie, Ingelin Steinsland, John M Hickey, and Gregor Gorjanc. 2019. "Flexible modelling of spatial variation in agricultural field trials with the R package INLA." *Theoretical and Applied Genetics* 132 (12): 3277-3293.

Serapiglia, Michelle J, Kimberly D Cameron, Arthur J Stipanovic, and Lawrence B Smart. 2012. "Correlations of expression of cell wall biosynthesis genes with variation in biomass composition in shrub willow (Salix spp.) biomass crops." *Tree genetics & genomes* 8 (4): 775-788.

Sharma, Deepika, Raakhi Gupta, and Ila Joshi. 2014. "Nutrient analysis of raw and processed soybean and development of value added soybean noodle." *Invent J* 1: 1-5.

Shen, Yuan, Tingting Lei, Xiaoyun Cui, Xiaoyun Liu, Shaoli Zhou, Yu Zheng, Florence Guérard, Emmanuelle Issakidis-Bourguet, and Dao-Xiu Zhou. 2019. "Arabidopsis histone deacetylase HDA 15 directly represses plant response to elevated ambient temperature." *The Plant Journal* 100 (5): 991-1006.

Sherasia, P. L., M. R. Garg, and B. M. Bhanderi. 2017. *Pulses and their by-products as animal feed.* Rome: Food and Agriculture Organization of the United Nations (FAO).

Shu, Kai, and Wenyu Yang. 2017. "E3 Ubiquitin Ligases: Ubiquitous Actors in Plant Development and Abiotic Stress Responses." *Plant and Cell Physiology* 58 (9): 1461-1476. https://doi.org/10.1093/pcp/pcx071.

Sierra, V., N. Aldai, P. Castro, K. Osoro, A. Coto-Montes, and M. Oliván. 2008. "Prediction of the fatty acid composition of beef by near infrared transmittance spectroscopy." *Meat Science* 78 (3): 248-255. https://doi.org/https://doi.org/10.1016/j.meatsci.2007.06.006.

Silva, Y, G Trasmonte, and L Giráldez. 2010. "Variabilidad de las Lluvias en el Valle del Mantaro Memoria del Subproyecto: Pronóstico Estacional de Lluvias Y Temperatura en la Cuenca del río Mantaro Para su Aplicación en la Agricultura." *Fondo Editorial CONAM-Instituto Geofísico del Perú: Lima, Peru*.

Sinjushin, Andrey A. 2015b. "Mutations of determinate growth and their application in legume breeding." *Legume Perspectives* 6: 14-15.

Slavov, Gancho, Gordon Allison, and Maurice Bosch. 2013. "Advances in the genetic dissection of plant cell walls: tools and resources available in Miscanthus." *Frontiers in Plant Science* 4. https://doi.org/10.3389/fpls.2013.00217.

Smil, Vaclav. 1999. "Crop Residues: Agriculture's Largest Harvest: Crop residues incorporate more than half of the world's agricultural phytomass." *Bioscience* 49 (4): 299-308.

Smith, Pete, Helmut Haberl, Alexander Popp, Karl-heinz Erb, Christian Lauk, Richard Harper, Francesco N Tubiello, Alexandre de Siqueira Pinto, Mostafa Jafari, and Saran Sohi. 2013. "How much land-based greenhouse gas mitigation can be achieved without

compromising food security and environmental goals?" *Global change biology* 19 (8): 2285-2302.

Smith, Stuart C., Rachel Choy, Stuart K. Johnson, Ramon S. Hall, Alida C.M. Wildeboer-Veloo, and Gjalt W. Welling. 2006. "Lupin kernel fiber consumption modifies fecal microbiota in healthy men as determined by rRNA gene fluorescent in situ hybridization." *European Journal of Nutrition* 45 (6): 335-341. https://doi.org/10.1007/s00394-006-0603-1.

Stawinski, S, and W Rybinski. 2001. "Domestication of Andean lupine by means of mutation." 4th European Conference on Grain Legumes, AEP, Kraków, Poland.

Stegmann, Paul, Marc Londo, and Martin Junginger. 2020. "The circular bioeconomy: Its elements and role in European bioeconomy clusters." *Resources, Conservation & Recycling: X* 6: 100029.

Stein, Ofer, and David Granot. 2019. "An overview of sucrose synthases in plants." *Frontiers in plant science* 10: 95.

Stewart, Jaclyn J, Takuya Akiyama, Clint Chapple, John Ralph, and Shawn D Mansfield. 2009. "The effects on lignin structure of overexpression of ferulate 5-hydroxylase in hybrid poplar1." *Plant physiology* 150 (2): 621-635.

Strozycki, Pawel M., Anna Szczurek, Barbara Lotocka, Marek Figlerowicz, and Andrzej B. Legocki. 2007. "Ferritins and nodulation in Lupinus luteus: iron management in indeterminate type nodules." *Journal of Experimental Botany* 58 (12): 3145-3153. https://doi.org/10.1093/jxb/erm152.

Sul, Jae Hoon, Lana S Martin, and Eleazar Eskin. 2018. "Population structure in genetic studies: Confounding factors and mixed models." *PLoS genetics* 14 (12): e1007309.

Sweetingham, M, J Clements, B Buirchell, S Sipsas, G Thomas, J Quealy, R Jones, C Francis, and CG Smith. 2006. "Preliminary breeding and development of Andean lupin for Australian agriculture." México, where old and new world lupins meet. Proceedings of the 11th International Lupin Conference,, 4-9 May 2005, Guadalajara, Jalisco, Mexico, 2006.

Swiecicki, WK, and C Nawrot. 2004. "Zasoby genowe rodzaju Lupinus." *Zeszyty Problemowe Postępów Nauk Rolniczych* 497 (2).

Talhinhas, P., J. Neves-Martins, and J. Leitao. 2003. "AFLP, ISSR and RAPD markers reveal high levels of genetic diversity among Lupinus spp." *Plant Breeding* 122 (6): 507-510. https://doi.org/10.1111/j.1439-0523.2003.00892.x.

Tallentire, CW, SG Mackenzie, and I Kyriazakis. 2018. "Can novel ingredients replace soybeans and reduce the environmental burdens of European livestock systems in the future?" *Journal of Cleaner Production* 187: 338-347.

Tapia, Mario E. 2015. "El tarwi, lupino Andino." *Tarwi, tauri o chocho*.

Tavares, E.Q., De Souza, A.P., Romim, G.H., Grandis, A., Plasencia, A., Gaiarsa, J.W., Grima-Pettenati, J., de Setta, N., Van Sluys, M.A. and Buckeridge, M.S., 2019. The control of endopolygalacturonase expression by the sugarcane RAV transcription factor during aerenchyma formation. *Journal of experimental botany*, *70*(2), pp.497-506.

Taylor, Candy M, Lars G Kamphuis, Weilu Zhang, Gagan Garg, Jens D Berger, Mahsa Mousavi-Derazmahalleh, Philipp E Bayer, David Edwards, Karam B Singh, and Wallace A Cowling. 2019. "INDEL variation in the regulatory region of the major flowering time gene LanFTc1 is associated with vernalization response and flowering time in narrow-leafed lupin (Lupinus angustifolius L.)." *Plant, cell & environment* 42 (1): 174-187.

Tian, Shuang-Qi, Ren-Yong Zhao, and Zhi-Cheng Chen. 2018. "Review of the pretreatment and bioconversion of lignocellulosic biomass from wheat straw materials." *Renewable and Sustainable Energy Reviews* 91: 483-489.

Tian, Zhixi, Xiaobo Wang, Rian Lee, Yinghui Li, James E Specht, Randall L Nelson, Phillip E McClean, Lijuan Qiu, and Jianxin Ma. 2010. "Artificial selection for determinate growth habit in soybean." *Proceedings of the National Academy of Sciences* 107 (19): 8563-8568.

Tiekstra, and Adriaanse. 2014. "Tomato packaging—packing tomatoes in a box made from tomato side streams, closed loop approach." PTS Symposium Innovative Packaging Munich, Germany, 2014.

Ţîţei, Victor. 2020. "Some biological features and biomass quality of Lupinus albus and Lupinus luteus in Moldova."

Torres, Andres F, Tim van der Weijde, Oene Dolstra, Richard GF Visser, and Luisa M Trindade. 2013. "Effect of maize biomass composition on the optimization of dilute-acid pretreatments and enzymatic saccharification." *Bioenergy Research* 6 (3): 1038-1051.

Trugo, L. C., D. von Baer, and E. von Baer. 2003. "LUPIN." In *Encyclopedia of Food Sciences and Nutrition (Second Edition)*, edited by Benjamin Caballero, 3623-3629. Oxford: Academic Press.

Turner, SD. 2018. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. J. Open Source Softw. 3: 731.

Udall, Joshua A, Pablo A Quijada, Hieronim Polewicz, Robert Vogelzang, and Thomas C Osborn. 2004. "Phenotypic effects of introgressing Chinese winter and resynthesized Brassica napus L. germplasm into hybrid spring canola." *Crop science* 44 (6): 1990-1996.

Vaknin, Yiftach, Miriam Ghanim, Shahar Samra, Levana Dvash, Efrat Hendelsman, Dan Eisikowitch, and Yael Samocha. 2011. "Predicting Jatropha curcas seed-oil content, oil composition and protein content using near-infrared spectroscopy—A quick and non-destructive method." *Industrial Crops and Products* 34 (1): 1029-1034.

van der Weijde, Tim, Oene Dolstra, Richard G. F. Visser, and Luisa M. Trindade. 2017. "Stability of Cell Wall Composition and Saccharification Efficiency in Miscanthus across Diverse Environments." *Frontiers in Plant Science* 7. https://doi.org/10.3389/fpls.2016.02004.

Van Krimpen, MM, P Bikker, IM Van der Meer, CMC Van der Peet-Schwering, and JM Vereijken. 2013. *Cultivation, processing and nutritional aspects for pigs and poultry of European protein sources as alternatives for imported soybean products.* Wageningen UR Livestock Research.

van Rossum, Bart-Jan, Willem Kruijer, Fred van Eeuwijk, Martin Boer, Marcos Malosetti, Daniela Bustos-Korts, Emilie Millet, Joao Paulo, M Verouden, and R Wehrens. 2020a. "Package 'statgenGWAS'." *R package version* 1 (7).

VanRaden, Paul M. 2008. "Efficient methods to compute genomic predictions." *Journal of dairy science* 91 (11): 4414-4423.

Varshney, Rajeev K, Abhishek Bohra, Jianming Yu, Andreas Graner, Qifa Zhang, and Mark E Sorrells. 2021. "Designing future crops: genomics-assisted breeding comes of age." *Trends in Plant Science* 26 (6): 631-649.

Velazco, Julio G, María Xosé Rodríguez-Álvarez, Martin P Boer, David R Jordan, Paul HC Eilers, Marcos Malosetti, and Fred A Van Eeuwijk. 2017. "Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model." *Theoretical and Applied Genetics* 130 (7): 1375-1392.

Vicente Rojas, Juan José. 2016. "El cultivo de Tarwi (Lupinus mutabilis Sweet) en el Estado Plurinacional de Bolivia." *Revista Científica de Investigación INFO-INIAF*: 88.

Vogel, B. 2014. "Marker assisted selection: A biotechnology for plant breeding without genetic engineering." *Smart breeding: The next generation. Greenpeace, Amsterdam, The Netherlands*.

Von Baer, E. 2011. "Domestication of Andean lupin (L. mutabilis)." Proc. 13th Int. Lupin Conf., Poznań.

Von Baer, E., and D. Von Baer. 1988. "Lupinus mutabilis: Cultivation and breeding." 1988.

Voora, Vivek, Cristina Larrea, and Steffany Bermudez. 2020. *Global market report: Soybeans*. JSTOR.

Vos, Peter G, M João Paulo, Roeland E Voorrips, Richard GF Visser, Herman J van Eck, and Fred A van Eeuwijk. 2017. "Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato." *Theoretical and Applied Genetics* 130 (1): 123-135.

Wahl, Vanessa, Luise H Brand, Ya-Long Guo, and Markus Schmid. 2010. "The FANTASTIC FOUR proteins influence shoot meristem size in Arabidopsis thaliana." *BMC Plant Biology* 10 (1): 1-12.

Wang, Shaofang, Steve Errington, and Heng How Yap. 2008. "Studies on carotenoids from lupin seeds." Lupins for Health and Wealth, Proceedings of the 12th International Lupin Conference, Fremantle, Australia.

Wang, H., Avci, U., Nakashima, J., Hahn, M.G., Chen, F. and Dixon, R.A., 2010. Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proceedings of the National Academy of Sciences*, *107*(51), pp.22338-22343.

Wang, Xinjing, Bojing Du, Meng Liu, Na Sun, and Xiaoting Qi. 2013."Arabidopsis transcription factor WRKY33 is involved in drought by directly regulating the expression of CesA8." *American Journal of Plant Sciences*, *4 (6A)*

Watson, Christine A, Moritz Reckling, Sara Preissel, Johann Bachinger, Göran Bergkvist, Tom Kuhlman, Kristina Lindström, Thomas Nemecek, Cairistiona FE Topp, and Aila Vanhatalo. 2017. "Grain legume production and use in European agricultural systems." *Advances in Agronomy* 144: 235-303.

Weigel, Detlef, and Ove Nilsson. 1995. "A developmental switch sufficient for flower initiation in diverse plants." *Nature* 377 (6549): 495-500.

Weissmann, E, and S Weissmann. 1992. "Variation of various constituents in seeds of Lupinus mutabilis under German climatic conditions." *Agrimed Research Programme-Lupinus mutabilis: its adaptation and production under European pedoclimatic conditions (Commission of the European Communities). EUR* 14102: 155-163.

Wickland, Daniel P, and Yoshie Hanzawa. 2015. "The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms." *Molecular Plant* 8 (7): 983-997.

Williams, Watkin. 1979. "Studies on the development of lupins for oil and protein." *Euphytica* 28 (2): 481-488. https://doi.org/10.1007/bf00056608.

Williams, Watkin, Jill E. M. Harrison, and Sumithra Jayasekera. 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, and L Witte. 1985. "Quinolizidine alkaloids as nitrogen source for lupin seedlings and cell cultures." *Zeitschrift für Naturforschung C* 40 (11-12): 767-775.

Wink, M. 1990. "Plant breeding: low or high alkaloid content." 1990.

Wink, Michael. 1993. "Allelochemical properties or the raison d'etre of alkaloids." In *The alkaloids: chemistry and pharmacology*, 1-118. Elsevier.

Wink, Michael, Carsten Meißner, and Ludger Witte. 1995. "Patterns of quinolizidine alkaloids in 56 species of the genus Lupinus." *Phytochemistry* 38 (1): 139-153. https://doi.org/https://doi.org/10.1016/0031-9422(95)91890-D.

Wink, Michael, Femando Merino, and Ernst Käss. 1999. "Molecular evolution of lupins (Leguminosae: Lupinus)." Lupin, an ancient crop for the New Millenium. Proceedings of the 9th International Lupin Conference. Klink/Müritz.

Wolko, B, and NF Weeden. 1990. "Isozyme number as an indicator of phylogeny in Lupinus." *Genetica Polonica* 31 (3-4): 179-187.

Wolko, Bogdan, Jon C. Clements, Barbara Naganowska, Matthew N. Nelson, and Hua'an Yang. 2011. "Lupinus." In *Wild Crop Relatives: Genomic and Breeding Resources: Legume Crops and Forages*, edited by Chittaranjan Kole, 153-206. Berlin, Heidelberg: Springer Berlin Heidelberg.

Xi, Wanyan, Chang Liu, Xingliang Hou, and Hao Yu. 2010. "MOTHER OF FT AND TFL1 regulates seed germination through a negative feedback loop modulating ABA signaling in Arabidopsis." *The Plant Cell* 22 (6): 1733-1748.

Yang, S Samuel, Wayne Wenzhong Xu, Mesfin Tesfaye, JoAnn FS Lamb, Hans-Joachim G Jung, Kathryn A VandenBosch, Carroll P Vance, and John W Gronwald. 2010. "Transcript profiling of two alfalfa genotypes with contrasting cell wall composition in stems using a cross-species platform: optimizing analysis by masking biased probes." *BMC genomics* 11 (1): 1-18.

Yao, D, ZZ Liu, J Zhang, SY Liu, J Qu, SY Guan, LD Pan, D Wang, JW Liu, and PW Wang. 2015. "Analysis of quantitative trait loci for main plant traits in soybean." *Genet Mol Res* 14 (2): 6101-6109.

Yazaki, Kazufumi, Akifumi Sugiyama, Masahiko Morita, and Nobukazu Shitan. 2008. "Secondary transport as an efficient membrane transport mechanism for plant secondary metabolites." *Phytochemistry Reviews* 7 (3): 513-524. https://doi.org/10.1007/s11101-007-9079-8

Yu, Si-in, Hyojin Kim, Dae-Jin Yun, Mi Chung Suh, and Byeong-ha Lee. 2019. "Post-translational and transcriptional regulation of phenylpropanoid biosynthesis pathway by Kelch repeat F-box protein SAGL1." *Plant molecular biology* 99 (1): 135-148.

Zamora, HDZ, ÁM Zamora-Burbano, LHR Varão, TAL Silva, JDC Medina, and D Pasquini. 2020. "Potencial do tremoço andino (Lupinus mutabilis) para produção de biodiesel via rota metílica: Uma revisão."

Zhang, Huawei, Feng Cui, Yaorong Wu, Lijuan Lou, Lijing Liu, Miaomiao Tian, Yuese Ning, Kai Shu, Sanyuan Tang, and Qi Xie. 2015. "The RING finger ubiquitin E3 ligase SDIR1 targets SDIR1-INTERACTING PROTEIN1 for degradation to modulate the salt stress response and ABA signaling in Arabidopsis." *The Plant Cell* 27 (1): 214-227.

Zhang, Xuebin, Mingyue Gou, and Chang-Jun Liu. 2013. "Arabidopsis Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase." *The Plant Cell* 25 (12): 4994-5010.

Zhang, Yan, Bo Zhang, Tongwen Yang, Jie Zhang, Bin Liu, Xiangqiang Zhan, and Yan Liang. 2020. "The GAMYB-like gene SlMYB33 mediates flowering and pollen development in tomato." *Horticulture research* 7.

Zhao, Xuebing, Lihua Zhang, and Dehua Liu. 2012. "Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose." *Biofuels, Bioproducts and Biorefining* 6 (4): 465-482.

Zhao, Ya Li, Abdughani Dolat, Yosef Steinberger, Xin Wang, Amarjan Osman, and Guang Hui Xie. 2009. "Biomass yield and changes in chemical composition of sweet sorghum cultivars grown for biofuel." *Field Crops Research* 111 (1-2): 55-64.

Zhao T., Holmer R., De Bruijn S., Angenent G. C., Van Den Burg H. A., Schranz M. E. 2017. Phylogenomic synteny network analysis of MADS-box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. *Plant Cell* 29 1278–1292. 10.1105/tpc.17.00312

Zielińska-Dawidziak, Magdalena. 2015. "Plant ferritin--a source of iron to prevent its deficiency." *Nutrients* 7 (2): 1184-1201. https://doi.org/10.3390/nu7021184.

Zoga, M., A. Pawelec, R. Galek, and E. Sawicka-Sienkiewicz. 2008. "Morphological, cytological and molecular characteristics of parents and interspecific hybrid (Lupinus mutabilis LM-13 × Lupinus albus sensu lato)." Canterbury.

Zoghlami, Aya, and Gabriel Paës. 2019. "Lignocellulosic biomass: understanding recalcitrance and predicting hydrolysis." *Frontiers in chemistry*: 874.

Zotova, Lyudmila, Akhylbek Kurishbayev, Satyvaldy Jatayev, Nikolay P Goncharov, Nazgul Shamambayeva, Azamat Kashapov, Arystan Nuralov, Ainur Otemissova, Sergey Sereda, and Vladimir Shvidchenko. 2019. "The General transcription repressor TaDr1 is co-expressed with TaVrn1 and TaFT1 in bread wheat under drought." *Frontiers in genetics* 10: 63.

Summary

In the last decade, the transition towards a sustainable bioeconomy and plant protein based diets has become a central point of European political agendas. This transition provides a viable strategy to mitigate societal and environmental consequences of natural resource depletion, environmental pollution and climate change. Novel production systems are needed that combine sustainability to higher productivity, aiming at producing both food and renewable biological resources while improving ecosystem health. The use of multipurpose crops, developed to produce more outputs with less inputs, can play a pivotal role in helping to achieve this goal. In particular, the introduction of new protein crops, where agricultural side streams can be valorised into added value products, is of special interest for Europe to break its current dependency on the import of protein sources such as soybean.

In this thesis we explore the potential of introducing *Lupinus mutabilis*, an Andean legume, as a multipurpose crop on European marginal land for the production of protein, oil and lignocellulosic biomass. This species is characterized by an oil and protein content similar to that of soybean and is highly valued for its adaptability to colder climates and low input agriculture on marginal land. However, to date, *L. mutabilis* represents an important indigenous crop only in the Andean region and lacks a consistent breeding history. From our review of the current knowledge on *L. mutabilis* (**Chapter 2**) the development of locally adapted genotypes, characterized by early maturity, a determinate growth habit and consistent yields emerges as crucial for the successful introduction of this crop in Europe. In the past, the scarce accessibility of germplasm material has been overcome with the repeated use of a limited set of genetic resources and of induced mutation, leading to a narrowing of the genetic basis. Nowadays, the combined use of germplasm resources and modern approaches to broaden the genetic basis could aid the introgression of desirable traits for specific environments into stable lines, making it possible to develop *L. mutabilis* into a valuable crop for Europe.

To this end, the primary aim of this study was to exploit the natural variation characterizing this species by studying for the first time an ample collection of *L. mutabilis* accessions, including both germplasm resources from the Andean region and breeding lines from Europe With this work, we sought to identify useful genetic resources, adaptation strategies and tools to breed *L. mutabilis* varieties for specific european agronomic conditions. We started by evaluating phenotypic diversity and agronomic performance of the assembled collection both in the native environment (Andean region) and over two cropping conditions in Europe, as a winter crop for the Mediterranean region and as summer crop for North-Central Europe (**Chapter 3**). Overall, the different cropping conditions had a large effect on the morphology of *L. mutabilis*. Two different morphotypes were observed in Mediterranean and North-Central European conditions, characterized respectively by high grain yield (3 t/ha) and high biomass yield (50 t/ha). The presence of large diversity in flowering time, yield and architecture related traits candidate this

panel as a good starting point to select useful genetic resources for specific environments, and pointed out the superior performance of Peruvian germplasm in European environments.

The possibility of using *L. mutabilis* agricultural residues as lignocellulosic feedstock was also explored here for the first time. From our investigation of biomass yield and quality (**Chapter 4**), *L. mutabilis* emerged as a valuable source of high-quality biomass, characterized by high cellulose yields comparable to the ones of other common biomass feedstocks such as miscanthus and sweet sorghum. The broad diversity in biomass composition and yield assessed across the panel, was exploited to study the genetic architecture of this traits and to develop molecular tools to aid the breeding of this crop through a genome wide association study (GWAS). For each trait assessed 3 to 12 QTL were identified, supporting the idea that the genetic basis of biomass quality traits is generally controlled by several genes with minor effects. Underlying these loci, important genes involved in cellulose and sucrose synthesis, monolignol biosynthesis and pectin degradation were identified, which could play a role in reducing biomass recalcitrance and ultimately increasing fermentable sugars yield.

The same approach was also used to investigate the genetic regulation of agronomic traits related to yield, flowering time and plant architecture. The GWAS for these traits yielded seven environment specific QTL linked to biomass yield, plant height, pods production and flowering time **(Chapter 5**). The found QTL harboured genes involved in the regulation of major pathways influencing flowering and indeterminate/determinate growth habit, but also genes acting on these pathways in response to hormonal and environmental cues. Two QTL stable across environments were identified on chromosome 8 for flowering time.

We concluded our characterization of this large panel, by assessing the diversity in protein and oil content, and oil composition of *L. mutabilis* seeds (**Chapter 6**). Near-infrared reflectance spectroscopy (NIRS) was used to develop predictive models for a high-throughput characterization of seed quality in the panel cultivated under Mediterranean conditions. Bolivian and Ecuadorian material emerged as characterized by the highest oil (~15% dm) and protein (~44% dm) concentrations in the seeds respectively. Furthermore, due to a fatty acid profile rich in oleic, linoleic, palmitic and stearic acids, and low in linolenic acid, *L. mutabilis* oil appears suitable both for edible and industrial purposes, and especially promising for biodiesel production.

Altogether, our study of *L. mutabilis* provided new insights and genetic tools for the development of this crop for Europe. The extensive phenotypic characterization of the panel, the identification of QTL related to key agronomic and biomass quality traits, and the development of predictive models for high-throughput characterization here reported, provide the tools to initiate large breeding programs of *L. mutabilis* in Europe. Breeding of locally adapted genotypes for Mediterranean and North-Central cropping conditions can be now assisted by the knowledge generated in this study, and contribute to the introduction of a superior alternative for protein, oil and lignocellulosic biomass in Europe.

Sommario

Nell'ultimo decennio, la transizione verso un modello di bioeconomia circolare e verso diete a base di fonti proteiche vegetali e piu sostenibili e' diventato uno dei temi principali nelle agende politiche Europee. Questa transizione si presenta come un' importante strategia per mitigare le consequenze sociali e ambientali dettate dall'esaurimento delle risorse naturali, dagli allarmanti livelli di inquinamento e dai cambiamenti climatici. Occorrono nuovi sistemi di produzione che uniscano sostenibilita' a una maggiore produttivita', con l'obiettivo di produrre sia cibo che risorse biologiche rinnovabili, migliorando al contempo la salute dell'ecosistema. L'uso di colture polivalenti, in grado di produrre piu output con meno input, puo' svolgere un ruolo fondamentale nel raggiungimento di questo obiettivo. In particolare, l'introduzione di nuove colture proteiche, i cui prodotti di scarto possano essere convertiti in prodotti industriali a valore aggiunto appare particolarmente importante per spezzare l'attuale dipendenza dell'Europa dall'importo di fonti proteiche come la soia, necessarie a soddisfare la domanda interna.

Questa tesi investiga la potenziale introduzione del *Lupinus mutabilis*, un legume Andino, come coltura polivalente su suoli marginali in Europa per la produzione di proteine, olio e biomasse. Questa specie e' caratterizzata da un alto contenuto proteico, simile a quello della soia, ed e' altamente apprezzata per la sua adattabilita' a climi freddi e ad una coltiivazione a basso input in zone marginali. Tuttavia, ad oggi, *L. mutabilis* rappresenta un'importante coltura solo nella regione andina e manca di una consistente storia di selezione. Dalla nostra revisione delle attuali conoscenze su *L. mutabilis* (**Capitolo 2**) lo sviluppo di genotipi adattati localmente che siano caratterizzati da maturità precoce, crescita determinata e rese costanti appare cruciale per garantire il successo dell'introduzione di questa coltura in Europa. In passato, la scarsa accessibilità del germoplasma è stata superata con l'uso ripetuto di limitate risorse genetiche e di mutazioni indotte, portando ad un restringimento della base genetica. Oggi, l'uso combinato di germoplasma e approcci moderni per ampliare la base genetica potrebbe favorire l'introgressione di tratti desiderabili per ambienti specifici in linee stabili, rendendo possibile lo sviluppo di *L. mutabilis* in una coltura di valore per l'Europa.

A tal fine, lo scopo principale di questo studio è stato quello di sfruttare la variazione naturale che caratterizza questa specie studiando per la prima volta un'ampia raccolta di accessioni di *L. mutabilis*, comprendente sia risorse di germoplasma proventienti dalla regione andina che linee selezionate in Europa. Con questo lavoro abbiamo cercato di identificare risorse genetiche utili, strategie di adattamento e strumenti per selezionare varietà di *L. mutabilis* per specifiche condizioni agronomiche in Europa. Abbiamo iniziato valutando la diversità fenotipica e le prestazioni agronomiche della collezione assemblata sia nell'ambiente nativo (regione andina) che in due possibili tipi di coltivazione in Europa, come coltura invernale per la regione mediterranea e come coltura estiva per l'Europa centro-settentrionale (**Capitolo 3**). Nel complesso, le diverse condizioni di coltivazione hanno avuto un forte effetto sulla morfologia di *L. mutabilis*. Nel Mediterraneo e nell'Europa centro-settentrionale sono stati osservati due differenti morfotipi,

caratterizzati rispettivamente da un'elevata resa in granella (3 t/ha) e da un'elevata resa in biomassa (50 t/ha). La presenza di un'ampia diversità tra le accessioni in tempi di fioritura, resa e tratti relativi all'architettura fa di questa collezione un buon punto di partenza per una selezione mirata di risorse genetiche per specifici ambienti e mette in evidenza le prestazioni superiori del materiale genetico andino in ambiente europeo.

In questo studio, esploriamo per la prima volta anche la possibilità di utilizzare i residui agricoli di *L. mutabilis* come biomassa lignocellulosica. Dalla nostra indagine sulla resa e sulla qualità della biomassa prodotta (**Capitolo 4**), *L. mutabilis* emerge come una preziosa fonte di biomassa residuale di alta qualità, caratterizzata da elevate rese di cellulosa paragonabili a quelle di comuni colture da biomassa come il miscanto e il sorgo dolce. L'ampia diversità in resa e qualità della biomassa riscontrata nella collezione è stata sfruttata, attraverso uno studio di associazione genomica (GWAS), per studiare l'architettura genetica di questi tratti e per sviluppare strumenti molecolari che possano accellerare lo sviluppo di varietà colturali per specifici ambienti. Per ognuno dei tratti studiati sono stati identificati da 3 a 12 QTL, a supporto dell'idea che la base genetica dei tratti legati alla qualità della biomassa è generalmente controllata da diversi geni con effetti minori. Alla base di questi loci sono stati identificati importanti geni coinvolti nella sintesi della cellulosa e del saccarosio, nella biosintesi del monolignolo e nella degradazione della pectina, che potrebbero svolgere un ruolo importante nel ridurre la recalcitranza della biomassa e nell'aumentare la resa in zuccheri fermentabili.

Lo stesso approccio è stato utilizzato per investigare la regolazione genetica dei tratti agronomici legati alla resa, al tempo di fioritura e all'architettura della pianta. Lo studio di associazione genomica (GWAS) per questi tratti ha evidenziato sette QTL per specifici ambienti legati alla resa in biomassa, all'altezza della pianta, al numero di baccelli prodotti e al tempo di fioritura (**Capitolo 5**). I QTL trovati contengono geni coinvolti nella regolazione dei principali meccanismi che determinano la fioritura e l'architettura della pianta ma anche geni che agiscono su questi meccanismi in risposta a stimoli ormonali e ambientali. Nei diversi ambienti, due QTL sul cromosoma 8 si sono rivelati stabilmente associati al tempo di fioritura.

Abbiamo concluso la caratterizzazione di questa ampia collezione di accessioni valutando la diversità in contenuto proteico e oleoso dei semi di *L. mutabilis* e nella composizione in acidi grassi dell'olio prodotto (**Capitolo 6**). La tecnica di spettroscopia nel vicino infrarosso (NIRS) è stata utilizzata per sviluppare modelli predittivi per uno screening high-throughput della qualità dei semi nella collezione coltivata in condizioni mediterranee. Le linee boliviane ed ecuadoriane si sono contraddistinte rispettivamente per le piu alte concentrazioni di olio (~15% dm) e proteine (~44% dm) nei semi. Inoltre, grazie a un profilo di acidi grassi ricco in acido oleico, linoleico, palmitico e stearico, ma povero in acido linolenico, l'olio di *L. mutabilis* appare adatto sia per scopi alimentari che industriali, e particolarmente promettente per la produzione di biodiesel.

Complessivamente, il nostro studio su *L. mutabilis* mette a disposizione nuova conoscenza e strumenti molecolari che possano favorire lo sviluppo di questa coltura in Europa. L'ampia caratterizzazione fenotipica della collezione, l'identificazione di QTL associati ad importanti tratti agronomici e di qualità della biomassa, e lo sviluppo di modelli predittivi per high-throughput screening qui riportati forniscono gli strumenti per avviare ampi programmi di selezione per *L. mutabilis* in Europa. La selezione di specifici genotipi per la coltivazione in ambiente Mediterraneo e in Europa centro-settentrionale può essere ora assistita dalle conoscenze generate in questo studio e contribuire all'introduzione di un'alternativa superiore per la produzione di proteine, olio e biomassa in Europa.

Acknowledgments

As all things come to an end, also this book and this journey had to end. And this was a long journey. A beautiful and tough one. Full of discovery, growth and formation. Made possible by people, support and connection. So, dear reader, thank you for reaching here and allow me to now take time to thank all the important people that took part in this journey actually and made it possible. It is a long list.

I would like to start with Luisa Trindade, the one that made all this happen, my promotor and supervisor. Thank you for believing and betting on me. Thank you for giving me the opportunity to work on this project and for allowing me - along the way - to become an independent scientist. I don't deny that being independent wasn't easy at first, but you gave me the possibility to explore, to make my own choices, to evaluate different paths, to make mistakes and most importantly, to grow. Thank you for always believing that this was going to happen, even when I didn't. For always smiling, through the ups and downs of this journey, supporting and trusting my work despite my own criticism and moments of despair.

I would like to proceed thanking my co-promotor, João Paulo, for her support in the past three years. I highly appreciate the total availability and cooperation you demonstrated in the last stretch of this journey. Thank you for the long talks, the encouraging words and all the statistical recommendations and explanations. Thank you also to Robert van Loo, for helping me when most needed. I am glad I could work with you and I remain extremely grateful for the patience and enthusiasm you put in sitting for hours next to me (or on the other side of the screen) to guide my first steps through codes and scripts.

The people that made this work possible are, however, many more. I would like to thank all the partners of the LIBBIO project for welcoming me on board and for the opportunity you gave me to see, touch, and taste what a European research project is. Travelling and meeting together to discuss our new thoughts and discoveries on Andean lupin always inspired me and made me feel part of the right community. A particular thanks goes to Prof. João Neves Martines and Sofia Alves for welcoming me and a huge field trial at the University of Lisbon, and to VandinterSemo and Gerard Bremen for helping me out with the field trials in the Netherlands. The weeks spent harvesting lupins in a field, be it under the hot sun of Portugal or the grey sky of the Netherlands, remain for me one of the dearest memory I hold from this PhD. I could of course not proceed further without first thanking the people from INIAP, for trusting us and giving us access to their genetic material, but also for maintaining a direct connection in the attempt of building common knowledge about this species. The efforts of Andres Torres, Maria de Lourdes Torres and Milton Gordillo Romero were certainly invaluable to foster cooperation between Europe and the Andean region. Finally, thank you to my students Tim, Eline and Wilrik for their willingness to learn, for being versatile in adapting to the limitations imposed by Covid and for

still contributing with a lot of work to this research. I learned a lot from supervising students, and I hope you also learned something from this experience.

Furthermore I would like to extend my thanks to all the past and present members of the BBE, a group that has hosted me as a family since the far away 2017. Thank you to Jordi, Behzad, Viviana and Andres for having welcomed me many years ago in the department and for making me feel at home. Thank you to Annamarie Dechesne, for opening to me the doors of the biochemical lab, for helping me out with all kinds of experiments and for always being willing to come up together with solutions. Thank you to Francesco, Kasper, Annick, Mites, Ornela, Mohamad, Hugo and all the other BBEs for sharing with me different pieces of this journey and for enriching it with an endless amount of scientific advice, complaints and laughs between the walls of our departments. Thank you especially to my office-mates for enduring my changes of mood and bearing both my hyper-enthusiastic and extra-complaining self, both quite noisy. Thank you Hugo and Jillis for your patience. Thank you Mohammad for having brought a whole new level of humor in our office, but also for being a very good friend to me.

Thank you life for Antonio and Antonino, the paranymphs you put on my path at the end of this long journey. Thank you for having recreated the magic union of the South of Italy, connecting in a perfect line Catania, Reggio and Francavilla Fontana. I got two amazing colleagues and office-mates, always up for a scientific confrontation, righteously critical about every result, extremely resourceful and helpful. But I also got two amazing friends, that in the last years have greatly supported me and helped me, with all possible means, to stay positive and finalize this journey. I am proud to share my stage with you, and even more proud to say that our journey does not end here.

Thank you also to all the others that have somehow collaborated to this work, and that probably I am forgetting about in this moment. Thanks to Dianka, Elma and Linda Kodde for their help in the molecular lab. Thanks to the people of Agros and Unifarm, and in particular to Wim van der Slikke, for helping with processing infinite amounts of lupin stems. Thank you to Nicole Trefflich and Danielle van der Wee for always willing to help out. Thank you to all the PhD students of PBR that were there with me 'in the struggle' these years, and in particular to Eleni. I am happy we shared this journey together, from Master to PhD.

Thanks to all the reviewers that, for free and in their free time, went through my manuscripts and provided valuable feedback becoming invisible mentors in my scientific path. Your help was very valuable and always highly appreciated. And thank you to you all, European tax payers. You might not be aware of it, but you are the ones really funding this research.

Thanks to my community, Droevendaal, where I was so lucky to live for the past seven years. This special bubble has changed me and shaped me into who I am today more than I could have ever imagined when I reached Wageningen. It has taught me what sharing and caring really means and, through the beautiful diversity of people that it hosts, it has constantly challenged me to discover something new, to look at things from a different perspective and to re-learn a lot from

scratch. Unfortunately, it has also taught me that we cannot always keep close to us our loved ones, hence during the past seven years I had to say goodbye many times. It has been heartbreaking, but today it fills my heart with joy to know that no matter how many kilometers separate us you always remain my beautiful community, scattered throughout the world. It fills my heart even more to know that many of you will be travelling those kilometers to be here today with me.

Thank you to my PhD models and dear friends: Jordi, Xabi and Ramon. Finally the moment has come in which I can be grateful to you for convincing me that starting this PhD was a great idea. Thank you for your guidance, friendship and trust.

Thank you to the old "Spanish mafia" and to all the other generations of beautiful people that followed: Mariana, Serena, Mara, Maura, Vero, Carmen, Marco, Emilio, Romanos, Max, Butturini, Miguel, Nina, Davide, Clea, Andrea, Claire, Fotis, Eliot and all the others that I am surely forgetting to mention here. I am grateful I got to meet you all.

Most importantly, thank you to the ones that took care of me on a daily basis, that laughed, cried and danced with me through the good and the bad of these years: thank you 105. Within the walls of this house I have met what I consider by now my multiple families: Noor, Tobi, Maria, Philippe, Fabian, Lucie, Femke, Julian, Viki, Arina, Teresa, Michele, Francesca, Alessandro, Carlota, Anna, Mario, Maarten. Thank you for what we built together, for all the effort we put in keeping the same spirit alive at every new round. The care and dedication towards each other that has always characterized 105 through the different families and years, makes me proud every day. Thank you Philippe and Tobi for being there from the beginning 'til the end.

Thank you to my beautiful women. I have been so lucky to encounter you, such beautiful and diverse examples of empowerment. Teresa, Maria, Arina, Serena, Noor, Viktoria, Mariana, Mara: you keep inspiring me every day.

Thank you Teresa, for the constant exchange, the full trust, the honesty and the complementarity characterizing our relationship. Non saro' mai in grado di spiegare a parole la magia del nostro incastro perfetto, quindi preferisco ringraziare quella meravigliosa congiunzione astrale che ha portato le nostre strade ad incrociarsi. To the moon and back.

Thank you Maria, for being like me. For sharing the same humor, attitude towards life and vision. For reading my mind, and still pretending to be surprised. La liamos alla donde vamos, and we are about to go far!

Grazie a Miriam, Filippo e Gabriele, coloro che sempre ci sono stati e sempre ci saranno. Siete i punti di riferimento che mi hanno permesso di arrivare fino a qui. Siete casa.

E per finire grazie alla mia splendida famiglia. Grazie a mio fratello, Giulio, per il suo amore incondizionato. Grazie ai miei genitori, Leonardo e Giovanna per avermi fatta come sono. Per avermi cresciuta con dei valori, per avermi sostenuto e garantito un'educazione, per avermi

guidato nel raggiungimento dei miei obiettivi, ma soprattutto per avermi sempre lasciato libera di scegliere.

Grazie a me stessa, per essere arrivata sin qui.

"It is good to have an end to journey towards; but it is the journey that matters, in the end."

About the author

Agata Gulisano was born on the 9th of June, 1992 in Giarre, a small town on the east coast of Sicily, in the south of Italy. In 2012 she started a Bachelor in Agricultural Science at the University of Catania, where she graduated with honors in 2015. During the Bachelor she also spent six months at the Cukurova University of Adana, in Turkey, as part of the Erasmus Programme.

Her interest in studying plant biology and genetics led her to pursue a Master's degree in Plant Science at the University of Wageningen, with specialization in Plant Breeding and Genetic Resources. Her Master thesis at the group of Biobased Economy, in the Plant Breeding department of Wageningen University, focused on the morphological analysis of hemp stems used for fiber production. Her work at the laboratory of biochemistry resulted in the development of a new protocol to estimate the ratio bast/shive in fiber hemp, that was published together with Petit et al. in the journal of Frontiers in Plant Science.

After the Master Thesis, she had the opportunity to start working on Lupinus mutabilis, a forgotten Andean legume with the potential of becoming a multipurpose crop for Europe. During her minor thesis, she started reviewing the status of this crop and investigating the possibility of using the agricultural residues of this crop as a source of lignocellulosic biomass. Her work resulted in a review of current knowledge and in the setup of a preliminary protocol for the analysis of cell wall composition in Lupinus mutabilis.

Soon after, in 2018 she started a PhD in the same department on the field of Plant breeding for Bio-based Economy under the supervision of Prof. Dr Luisa Trindade. This PhD was granted within the European project LIBBIO (Lupinus mutabilis for increased biomass from marginal lands and value for biorefineries) and funded by the European Union and the Bio-based joint undertaking partnership. Her focus was on the study of genetic and phenotypic diversity present in a collection of *L. mutabilis* genetic resources, to aid the breeding of targeted lines for the cultivation of this crop in different European conditions. The results of her PhD project are described in this thesis.

List of publications

Petit, J., **Gulisano, A.,** Dechesne, A., & Trindade, L. M. (2019). Phenotypic variation of cell wall composition and stem morphology in hemp (Cannabis sativa L.): optimization of methods. *Frontiers in plant science*, *10*, 959.

Gulisano, A., Alves, S., Martins, J. N., & Trindade, L. M. (2019). Genetics and breeding of Lupinus mutabilis: An emerging protein crop. *Frontiers in Plant Science*, *10*, 1385.

Gulisano, A., Alves, S., Rodriguez, D., Murillo, A., Van Dinter, B. J., Torres, A. F., ... & Trindade, L. M. (2022). Diversity and agronomic performance of Lupinus mutabilis germplasm in European and Andean environments. *Frontiers in Plant Science*, *13*.

Gulisano, A., Dechesne, A., Paulo, M. J., & Trindade, L. M (2022). Investigating the potential of Andean lupin as a lignocellulosic feedstock for Europe: first genome-wide association study on L. mutabilis biomass quality. *GCB Bioenergy*.

Gulisano, A., Lippolis, A., Van Loo, E. N., Paulo, M. J., & Trindade, L. M (2022). Genome wide association study of agronomic traits in Andean lupin. *Submitted.*

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to: **Agata Gulisano**
Date: **09 December 2022**
Group: **Plant Breeding - Biobased Economy**
University: **Wageningen University & Research**

1) Start-Up Phase	*date*
► **First presentation of your project**	
"Genomic data resources for L. mutabilis breeding in Europe", Biobased Economy Group meeting	22 May 2018
► **Writing or rewriting a project proposal**	
Investigating the genetic variability of L. mutabilis using Genome Wide Association mapping (GWAS).	31 May 2018
► **Writing a review or book chapter**	
Gulisano A, Alves S, Martins JN,Trindade LM. Genetic and breeding of Lupinus mutabilis: an emerging protein crop. *Frontiers in Plant Science*. 2019 doi:0.3389/fpls.2019.01385	30 Oct 2019
► **MSc courses**	
Subtotal Start-Up Phase	*13.5**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD student days 'Get2Gether', Soest, NL	15-16 Feb 2018
EPS PhD student days 'Get2Gether', Soest, NL	11-12 Feb 2019
EPS PhD student days 'Get2Gether', Soest, NL	10-11 Feb 2020
EPS PhD student days 'Get2Gether', online	01-02 Feb 2021
► **EPS theme symposia**	
EPS theme 4 'Genome biology', Wageningen, NL	13 Dec 2019
EPS theme 3 'Metabolism and adaptations', online	30 Oct 2020
EPS theme 2 'Interactions between plants and biotic agents', online	09 Feb 2021
► **Lunteren Days and other national platforms**	
Annual meeting 'Experimental Plant Science', Lunteren, NL	9-10 Apr 2018
Annual meeting 'Experimental Plant Science', Lunteren, NL	12-13 Apr 2021
► **Seminars (series), workshops and symposia**	
Seminar: "Metabolic engineering to enhance photosynthesis and increase crop yield" , Dr. Andrew J. Simkin,Wageningen, NL	21 Mar 2019
Seminar: "Development of genetic tools in miscanthus breeding", Prof. Toshihiko Yamada, Wageningen, NL	05 Sep 2019
Seminar: "Genome Editing - A "precision breeding to fork" strategy", Plant ETP, online	25 Nov 2020
Seminar: "Challenges as a PhD and postdoc in Science and tips to overcome those", Dr. Stefan Geisen, Wageningen, NL	20 Jan 2021
Seminar: "Plants ad alternaive protein sources", Dr. P. Annicchiarico, Prof. N. Sozer, Ass.Prof. S. U. Andersen, EPSO, online	20 Jan 2022
The Carpentries workshop: "Genomic Data", Wageningen, NL	05-06 Feb 2019
EPS & PE&RC workshop: "Breeding for diversity, Opportunities and Challenges", Wageningen, NL	30 Oct 2019
Symposium: "CRISPR-Cas - from evolution to revolution", Prof. E. Koonin, Prof. J. van der Oost, Prof. N. Geijsen, A. Bredenoord, Wageningen, NL	08 Mar 2018
Symposium: "Polyploidy genetics and breeding", C. Malliepaard, C. Hackett, J. Endelman, A. Post, P. Bourke, Wageningen, NL	14 Jun 2018
► **Seminar plus**	
► **International symposia and congresses**	
Conference: "LIBBIO: New alternatives in Lupin agronomy and processing", Lisbon, Portugal	18-19 Apr 2018
Conference: "LIBBIO consortium meeting", Iași, RO	26-27 Sept 2018

Conference: "Plant proteins: Alternative sources and their applications in food", at DIL, Quakenbrück, DE	11 Apr 2019
Conference: "Canapa Mundi, 4th Annual Conference on Hemp", Rome, IT	21-23 Feb 2020
Conference: "National Conference on Therapeutic Cannabis: Re-discovering the multiple uses of hemp in the third millenia", online	06-07 Feb 2021
► **Presentations**	
Poster: "Lupinus mutabilis: an emerging crop for Europe", Annual meeting 'Experimental Plant Science' 2021, Lunteren, NL	12-13 Apr 2021
Talk: "Progress on biomass quality phenotyping methods", LIBBIO meeting, Iaşi, RO	26 Sep 2018
Talk: "Lupinus Mutabilis. Using natural diversity as breeding tool", Plant Breeding Monday Seminars, Wageningen, NL	29 Oct 2018
Talk: "Phenotypic diversity of L. mutabilis in Europe", Plant Breeding Monday Seminars, Wageningen, NL	4 May 2020
Talk: "Hemp for fiber production: phenotypic variation of cell wall composition and stem morphology", Canapa Mundi, 4th Annual Conference on Hemp, Rome, IT	22 Feb 2020
Talk: "Hemp for fiber production", National conference on Therapeutic Cannabis, online	6 Feb 2021
Talk: "Lupinus mutabilis: exploiting diversity to guide breeding", CREA (Council for Agricultural research), Lodi, IT	21 Sep 2022
► **IAB interview**	
► **Excursions**	
EPS company visit: Koppert Biological Systems	26 Oct 2018
Subtotal Scientific Exposure	*15.5**

3) In-Depth Studies	*date*
► **Advanced scientific courses & workshops**	
AMC-WUR course: "Electron Microscopy - The basics: from Amsterdam to Wageningen", Amsterdam & Wageningen, NL	12-16 Nov 2018
EPS-VLAG course: "Microscopy and Spectroscopy in Food and Plant Science", Wageningen, NL	7-9 May 2019
SLU-WUR course: "Plant Breeding and Biotechnology", Wageningen, NL	11-13 Jun 2019
ELIXIR-EPS course: "Gentle hands-on introduction to Python programming", online	02-03 Jul 2020
WIAS course: "Genomic prediction in animal and plant breeding, Wageningen, NL	17-21 Oct 2022
► **Journal club**	
► **Individual research training**	
Subtotal In-Depth Studies	*6.2**

4) Personal Development	*date*
► **General skill training courses**	
WGS course: "Project and Time management", Wageningen, NL	8 Nov - 18 Dec 2018
PhD discussion club Plant Breeding (peer consultation), Wageningen, NL	Jan-Dec 2019
YoungWUR workshop: "How to create Impactful Infographics and Data Visuals", David MacCandless, online	24 Nov 2020
EPS Writing Support Group, online	Jan - Mar 2021
EPS course: "How to make a movie about your research", online	02-03 Mar 2021
ELSS course: "Navigating Brussels with ELLS 2022 - How can the EU help you to develop your scientific career", Brussels, BE	31 May- 2 Jun 2022
► **Organisation of meetings, PhD courses or outreach activities**	
Organiser and host of the weekly 'Plant Breeding Monday seminars'	Jan-Dec 2021
► **Membership of EPS PhD Council**	
Subtotal Personal Development	*7.3**

TOTAL NUMBER OF CREDIT POINTS*	**42.5**

** A credit represents a normative study load of 28 hours of study.*

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits.

The presented research has received funding from the Bio-based Industries Joint Undertaking private-public partnership under the European Union's Horizon 2020 research and innovation program under grant agreement No 720726 for the LIBBIO (Lupinus mutabilis for increased biomass from marginal lands and value for biorefineries) project.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover concept and design: Agata Gulisano

Printed by: ProefschriftMaken.nl