



**HAL**  
open science

## Proteomics for exploiting diversity of lupin seed storage proteins and their use as nutraceuticals for health and welfare

Francisco Cabello-Hurtado, Jean Keller, José Ley, Rosa Sanchez-Lucas, Jesús V. Jorrín-Novo, Abdelkader Aïnouche

### ► To cite this version:

Francisco Cabello-Hurtado, Jean Keller, José Ley, Rosa Sanchez-Lucas, Jesús V. Jorrín-Novo, et al.. Proteomics for exploiting diversity of lupin seed storage proteins and their use as nutraceuticals for health and welfare. *Journal of Proteomics*, 2016, Food and Crop Proteomics, 143, pp.57-68. 10.1016/j.jprot.2016.03.026 . hal-01295645

**HAL Id: hal-01295645**

**<https://univ-rennes.hal.science/hal-01295645>**

Submitted on 4 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Proteomics for exploiting diversity of lupin seed storage proteins and their use as nutraceuticals for health and welfare

Francisco Cabello-Hurtado<sup>1, \*</sup>, Jean Keller<sup>1</sup>, José Ley<sup>1</sup>, Rosa Sanchez-Lucas<sup>2</sup>, Jesús V. Jorrín-Novo<sup>2</sup>, Abdelkader Aïnouche<sup>1</sup>

<sup>1</sup>UMR CNRS 6553 Ecobio, OSUR (Observatoire des Sciences de l'Univers de Rennes), University of Rennes 1, 263 av. du Général Leclerc, 35042 Rennes, France.

<sup>2</sup>Agroforestry and Plant Biochemistry and Proteomics Research Group, Dpt. Biochemistry and Molecular Biology, University of Cordoba-CeiA3, Cordoba, Spain.

\* Corresponding author: phone: +33223235022; fax: +33223235026; e-mail: francisco.cabello@univ-rennes1.fr

## Abstract

Lupins have a variety of both traditional and modern uses. In the last decade, reports assessing the benefits of lupin seed proteins have proliferated and, nowadays, the pharmaceutical industry is interested in lupin proteins for human health. Modern genomics and proteomics have hugely contributed to describing the diversity of lupin storage genes and, above all, proteins. Most of these studies have been centered on few edible lupin species. However, *Lupinus* genus comprises hundreds of species spread throughout the Old and New Worlds, and these resources have been scarcely explored and exploited. We present here a detailed review of the literature on the potential of lupin seed proteins as nutraceuticals, and the use of -omic tools to analyze seed storage polypeptides in main edible lupins and their diversity at the *Lupinus* inter- and intra-species level. In this sense, proteomics, more than any other, has been a key approach. Proteomics has shown that lupin seed protein diversity, where post-translational modifications yield a large number of peptide variants with a potential concern in bioactivity, goes far beyond gene diversity. The future extended use of second and third generation proteomics should definitely help to go deeper into coverage and characterization of lupin seed proteome.

**Keywords:** Conglutins; Nutraceuticals; Seed proteome; Lupin proteome; *Lupinus* germplasm.

## Contents

- I. Introduction
  2. Insight on lupin seed storage proteins
  3. Research areas in nutraceuticals for lupin seed proteins
    - 3.1. Benefits on cardiovascular disease related risks
    - 3.2. Benefits on diabetes and hyperglycemia
  4. Lupin diversity and genomic resources in lupin seed storage proteins
  5. Seed proteomics of lupin domestic species
  6. New directions on lupin seed proteomics research
  7. Conclusions
- Acknowledgements
- References

## 1. Introduction

Lupin is the common name for *Lupinus*, a genus with more than 200 species which belongs to the genistoid clade, one of the six main clades of the Fabaceae family [1]. Lupins are distributed over two main geographical centers of diversity: one is the Mediterranean basin and North Equatorial Africa (so-called Old World lupins), and the other extends through North and South America (so-called New World lupins). Lupins are minor legume crops characterized by their high seed protein content, and they are cultivated worldwide across more than one million hectares presenting a wide array of human uses [2-4].

Lupins have been historically and traditionally used as food for populations in both the Mediterranean and the Andean cultures. Traditional lupin seeds were and still are consumed after being boiled or abundantly washed with water, as a means to remove anti-nutritional seed alkaloids [5]. Medicinal uses of lupins (probably *Lupinus albus*) date all the way back to Hippocrates of Cos (400-356 BC). Related to the same period, there are numerous reports of their use as soil improving crops, namely in rotation with wheat [6]. It is not surprising that domesticated forms of two lupin species, *L. albus* (white lupin) and *L. mutabilis* (pearl lupin), emerged as the result of human cultivation for some 3000 years in the Mediterranean and Andean regions, respectively [6,7]. During the 19th century, *L. luteus* (yellow lupin) and *L. angustifolius* (narrow-leafed lupin), two other Mediterranean species, became fully domesticated crop plants [6,8]. Currently, these four species are the main edible *Lupinus* species. *L. albus* and *L. luteus* are mainly cultivated in Europe,

*L. angustifolius* is now mostly grown in Australia, and *L. mutabilis* is a common crop in South America. Average annual world production (2011-2013) exceeds a million tons, with Australia being the most important producer, followed by Poland, Chile, Ukraine and Belarus [9]. Lupins are mainly cultivated for their use as ruminant food and as green fertilizers but, even though still a minor activity, lupin production is increasingly being used for human nutrition and the interests of food and medical industries in lupin compounds have risen remarkably [2,6,10].

In modern societies, consumers are increasingly aware of the role of nutrition and diet in determining both human health and their state of wellbeing. A wide range of legume seeds is available for functional food production, among which lupin seeds stand out because of their unique chemical composition and biological functions. Thus, lupin seeds are very rich in proteins and dietary fiber, are almost free of starch and phytoestrogens, and contain numerous essential amino acids as well as minerals, antioxidants (e.g. tocopherol and polyphenols) and excellent lipids rich in unsaturated fatty acids [2,11,12]. Nowadays, lupin flour is used in pasta, bread, cakes or pastries [6], but the use of lupins is diversifying and expanding from traditional foods and recipes into new, functional foods. Paraphrasing the consensus document proposed by the International Life Sciences Institute (ILSI) Europe [13], food is regarded as “functional” when it is satisfactorily demonstrated to have a beneficial effect on at least one physiological function in the body, beyond its nutritional effects, in such a way that it contributes to either an improved state of health and well-being or to a reduced risk of disease,

Currently, there is an intriguing issue concerning the role of dietary proteins in the human body beyond their nutritional importance as a source of amino acids. As a matter of fact, certain whole proteins, or sometimes large fragments of them, are resistant to proteolysis in the gastro-intestinal tract [14], and in some cases are known to be beneficial to human health. This is the case of some entire proteins or peptides of a plant origin, such as the Bowman-Birk protease inhibitors, the soybean beta-conglycinin subunit, as well as an array of alpha-amylase inhibitors and lectins [15]. Lupin seed proteins are also presumed to release some bioactive peptides playing positive roles in fighting obesity, diabetes and cardiovascular diseases, as discussed in section 3. However, these studies have been mainly restricted to lupin flour or seed proteins from some varieties of *L. albus* and *L. angustifolius*. The huge diversity of the *Lupinus* genus is thus still widely unexplored and unexploited regarding their potential for health and welfare, which goes hand-in-hand with the poor coverage of *Lupinus* germplasm diversity in terms of seed proteins. Indeed, physiological and health-promoting effects are highly impacted by protein structure [16] and the analysis of the diversity of lupin seed proteins, in relation to their bioactive properties, will provide a deeper understanding of relationships between structure and bioactive properties aimed at the improvement

of lupin seed nutraceutical potential. This review focuses on lupin seed proteins with potential health promoting effects, and also has the purpose of outlining the use of conventional and upcoming proteomic technologies to unravel and exploit their diversity. The analyses of molecular data provided by this review are intended to be useful to field agronomists, food technologists, and nutritionists for a better exploitation of the potential of lupin seed proteins.

## 2. Insight on lupin seed storage proteins

Lupins have an average gross protein content of 33 to 40% of dry weight. This is considered to be a higher concentration than that found in other legumes, and is only comparable to the concentrations found in soybean [5,11]. Most of these proteins are located in cotyledonary storage vacuoles and correspond to seed storage proteins that will be used as a source of amino acids for initial seedling development. Flour produced from lupin seeds has been employed for the production of good quality protein isolates with excellent functional and nutritional properties [4]. Indeed, the use of this plant can be extended to the production of protein-concentrates, which, when added to other food products, can enrich their nutritional values, thus producing functional food [10]. Methods for processing raw legume material into protein-rich concentrates, or even isolates, are already available. Moreover, for lupin and pea this has proved to be economically feasible [17].

According to Osborne's classification [18], the protein pool of lupin seeds is divided into two large fractions: namely globulins and albumins. Albumins are soluble in water and, for the most part, have an enzymatic activity, although a storage role has also been attributed to some of them. Globulins, which are insoluble in water but soluble in salt solutions, outnumber albumins and constitute seed storage proteins in the strict sense. In addition, seed storage proteins have been classified into three classes according to their sedimentation coefficients, i.e. 11S, 7S and 2S classes. 11S and 7S are globulins. In the case of 2S storage proteins, even if they are in general most often termed albumins, in different species such in lupins they include proteins that has been classified either as albumins or globulins [11,19,20,22-25]. Main lupin seed proteins have been classified according to their electrophoretic mobility since the pioneer work of Blagrove and Gillespie [25] and Cerletti et al. [26], resulting in four groups: the two major storage proteins, called  $\alpha$ -conglutin and  $\beta$ -conglutin, and two minor groups, called  $\gamma$ -conglutin and  $\delta$ -conglutin (Table 1). Besides classifications based on physicochemical properties which have a certain empirical degree, the current availability of gene and protein sequences allow an unequivocal classification of lupin conglutins into these four main groups, and so in this review we will refer to the lupin seed storage

proteins as  $\alpha$ -,  $\beta$ -,  $\gamma$ - and  $\delta$ -conglutins. Later works in the early nineteen eighties analyzed conglutin isoelectric points [27], reporting the acidic nature of  $\alpha$ -,  $\beta$ - and  $\delta$ -conglutin and the basic nature of  $\gamma$ -conglutin, and the already clearly stated conglutin heterogeneity [27-30]. These former studies also reported the relative richness in sulfur-containing amino acids (cysteine, methionine) of  $\alpha$ -,  $\gamma$ - and  $\delta$ -conglutin and the presence in their monomeric structure of disulfide bonds, which are absent in  $\beta$ -conglutin, as well as the presence of carbohydrates, mostly neutral sugars, in most of the conglutins [21,27]. Since then, technical improvements in protein separation and analysis have allowed further characterization of conglutin chemical properties, structure (e.g. subunit composition and oligomeric status) and post-translational modifications (PTMs). Most of these studies have been focused on white and narrow-leafed lupins ([22,31], and others below), and only a few were conducted on the other edible lupin, such as *L. luteus* [32] and *L. mutabilis* [7]. Main features of lupin conglutin genes and proteins are summarized in Table 1.

$\alpha$ -Conglutin, also called legumin-like or 11S globulin family, is a major globulin which may account for amounts of about 35-37% of the total seed storage proteins in white lupin seeds [21]. In its mature form, after pro-polypeptide post-translational proteolysis,  $\alpha$ -conglutin is a hexameric protein. Each monomeric unit of  $\alpha$ -conglutin is composed of acidic, also called  $\alpha$ , and basic, also called  $\beta$ , subunits with molecular weight (MW) of between 42-52 kDa and of between 20-22 kDa, respectively. Both subunits are obtained by proteolytic cleavage of pro-monomer and remain linked by one disulphide bond [33]. White lupin  $\alpha$ -conglutin, unlike most legumin-like globulins, can be glycosylated in some of its  $\alpha$ -chains, those where a serine residue in a sequence N-X-S (where X is not Pro) has been shown to be present at the N-terminus [33,34]. However, this feature is not present in any of the *L. angustifolius*  $\alpha$ -conglutin sequences (Table 1) or in the only  $\alpha$ -conglutin sequence available in databases for *L. albus* (UniProtKB ID: Q53I54).

The most abundant lupin globulin is  $\beta$ -conglutin, also called vicilin-like or 7S globulin family, which represents around 44-45% of white lupin seed storage proteins [21]. They are relevant in human nutrition since they have been hypothesized as being the hypocholesterolemic component of lupin proteins for their high homology with the  $\alpha'$  subunit of  $\beta$ -conglycinin [15], the major bioactive component of soy proteins [35]. In its mature form after proteolytic processing,  $\beta$ -conglutin is a trimeric protein, each monomer being formed by three polypeptides of a low (17-20 kDa), medium (25-46 kDa) and high (53-64 kDa) molecular weight [36]. A  $\beta$ -conglutin precursor (54-75 kDa) constitutes each precursor monomer and is at the origin, through proteolysis, of the 3 polypeptides composing the monomer [36]. Likewise, many of the different  $\beta$ -conglutin polypeptides are glycosylated [37,38]. As already mentioned, no disulphide bridges are present in  $\beta$ -conglutin.

Among the conglutins, the  $\gamma$ -conglutin (7S globulin family) is also regarded with particular interest, since it displays some of the bioactive properties of lupins, claimed by pharmacopeias since ancient times and reported by recent research, which will be detailed later on in this review. The content of  $\gamma$ -conglutin in mature white lupin seeds accounts for about 5% of the total amount of proteins [21]. At a neutral pH,  $\gamma$ -conglutin monomer is a tetrameric protein which at acidic pH dissociates into monomers with a mass of about 50 kDa and is composed of two subunits (around 17 and 29 kDa), linked by disulfide bonds and obtained by proteolysis of a pro-polypeptide [37, 39]. Only the large subunit is glycosylated [40]. Even though conserved sequences within  $\gamma$ -conglutin suggest a xylanase-inhibiting activity, it lacks any kind of inhibitory activity against plant cell-wall degradation enzymes and its biological function still remains unknown [41].  $\gamma$ -Conglutin is quite unique since it does not degrade during seed germination and it is resistant to digestion [41]. A similar protein, called BG7S, has been discovered in soybean, and both BG7S and lupin  $\gamma$ -conglutin have been seen to bind insulin *in vitro* [42,43]. The quaternary structure of *L. angustifolius*  $\gamma$ -conglutin has recently been reported (UniProtKB - Q42369), which will help to explain the structural basis of its physiological role [44]. Two complete deduced amino acid sequences of  $\gamma$ -conglutin can be found in the the UniProt Knowledgebase (UniProtKB) for *L. albus* (Q9FEX1 and Q9FSH9) and for *L. angustifolius* (Q42369 and F5B8W7), and in all the sequences a glycosylation consensus one is present and resides in a cysteine-rich stretch in proximity to the N-terminal region (Table 1).

Finally,  $\delta$ -conglutin, belonging to 2S storage protein family, is the least abundant lupin conglutin, representing only around 3-4% of total conglutins in white lupin [45].  $\delta$ -Conglutin is a small monomeric protein (precursor of around 17 kDa), consisting of one small and one large subunit of around 4 and 9 kDa, respectively, which are linked by two disulfide bonds and are not glycosylated. In addition, the large subunit contains two additional intrachain disulfide bonds [23].

### 3. Research areas in nutraceuticals for lupin seed proteins

Studies exploring the bioactive properties and nutraceutical potential of lupin seed proteins has been performed using a diet implemented either with lupin flour or different fractions of lupin conglutins, the latter allowing one to focus on bioactive candidates. Three main isolation protocols have been implemented for conglutin fractionation, starting from defatted lupin flour. These fractionation protocols vary in their production scale and the purity of their fractions, these two parameters being inversely related. The “lab scale” procedure allows the purifying of the four

different conglutin families, and is based on saline solution (0.5 M NaCl) extraction at neutral/slightly alkaline pH, and further separation by different coupled chromatographic systems [30,43,46]. At the other end, the “industrial scale” procedure is based on extraction under alkaline conditions and further separation by isoelectric precipitation at pH 4-5, and it generates partially purified fractions, one containing  $\alpha$ -/ $\beta$ -/ $\delta$ -conglutin (precipitated fraction=LPI-E, with emulsifying properties) and the other  $\gamma$ -conglutin (soluble fraction=LPI-F, with foaming properties) [47]. An “intermediate scale” procedure, versatile and useful from lab to industrial scale, gives three fractions ( $\alpha$ -/ $\beta$ -conglutin,  $\delta$ -conglutin, and  $\gamma$ -conglutin fractions) of a greater purity compared to the industrial scale procedure. The latter procedure is based, like the industrial scale one, on extraction under alkaline conditions and separation by isoelectric precipitation at pH 4-5, but with further precipitation steps for increased purification [45].

The major scientific evidence of the utility of lupin seed proteins as a nutraceutical to help reduce cardiovascular disease and diabetes, as well as the factors playing a critical role, are discussed below.

### *3.1. Benefits on cardiovascular disease-related risks*

Cardiovascular disease (CVD) is the first cause of death in developed countries, but 80% of CVD deaths occur in developing countries. Therefore, we face a global epidemic that affects men and women equally, regardless of socio-economic barriers [48,49]. Many aspects of a person’s lifestyle influence the appearance of CVD besides an array of genetic predispositions, yet the main risk which can contribute to CVD is unhealthy diet habits. In economic terms, CVD treatment accounted for more than \$500 billion in the United States in 2010, considering drugs, surgery and revascularization procedures [50]. This is why prevention techniques should be encouraged, in particular upstream processes that, like diet, could improve cardiovascular health.

Studies in the late 20th century showed that numerous legumes have a beneficial hypocholesterolemic effect, even if they have only been proved useful when consumed by hypercholesterolemic subjects. One outcome of these results was, for example, that in 1999 the Food and Drug Administration approved a health claim for the cholesterol-lowering potential of soy proteins [51]. In this context, the interest in other legumes, e.g. lupins, has increased for the treatment and prevention of hypercholesterolemia, and new discoveries have been made in the last decade in lupin proteins. Thus, the same way that soy did [35], first studies using animal models showed that lupin seeds also constituted a potential dietary protein source, which could decrease



cardiovascular disease risk through a significant cholesterol-lowering effect and an anti-atherosclerotic activity in rats [52] and rabbits [53].

Since then, research on the hypocholesterolemic effect of lupin proteins has mainly concerned animal model trials. Thus, Bettzieche et al. [54] reported that total protein extracts from three different cultivars of narrow-leafed lupin (Boregine, Probor and Vitabor), and  $\alpha$ - $\beta$ -conglutin from cv Boregine, showed differential effects on the plasma lipids in rats fed on a hypercholesterolemic diet containing these lupin proteins (50 g/kg). Seed proteins from the cv Vitabor presented the most beneficial effects, followed by  $\alpha$ - $\beta$ -conglutin, and they were presumed to have been caused by the down-regulation of genes involved in lipid synthesis. The same group later reported that a diet enriched with narrow-leafed lupin proteins (200g/kg) showed a decreased triglyceride and cholesterol concentration in plasma, and increased triglycerides in liver and milk of hypercholesterolemic growing mouse dams [55]. However, the offspring of dams fed with lupin proteins gained less weight than the control group. This fact, together with the triglyceride-raising effect in plasma, led Bettzieche et al. [55] to advise a limited consumption of lupin protein during pregnancy and lactation. The hypocholesterolemic effect of lupin seed proteins was also reported by both Parolini and co-authors [56] using rats fed with a diet enriched with *L. angustifolius* seed proteins, and Fontanari and co-authors [57] using hypercholesterolemic/hyperlipidemic hamsters fed with a diet enriched with *L. albus* seed proteins. The latter study also showed a beneficial effect on blood triglycerides levels and fat accumulation in liver. Kapravelou et al. [58] later supported such claims with a similar study, but this time in rats, and they further noted a diminution in blood glucose levels. Recently, Radtke et al. [59] also confirmed that *L. angustifolius*  $\gamma$ -conglutin (200 g/kg) exerts a beneficial impact on serum cholesterol in apoE-deficient mice, a model of atherosclerosis, but this did not result in any obvious atheroprotective effect. The only study made in other animal models apart from rodents, was the one conducted by Radtke et al. [60] using pigs fed with a chow containing either a *L. angustifolius* protein isolate or casein as a control. The lupin protein reduced plasma cholesterol concentration and increased the cholesterol output in feces. They also confirmed that two intestine sterol transporters (NPC1L1 and SR-BI), which are involved in the transport of cholesterol from the intestinal lumen into the intestinal epithelial cells, were significantly down-regulated and that this decrease was partially responsible for the cholesterol-lowering effects.

On the other hand, most of the studies performed in humans evaluated the effects of lupin fiber or lupin flour, and only a few studies carried out human trials to test the hypocholesterolemic effect of lupin seed proteins. In a randomized, double-blind, placebo-controlled study performed in 43 moderately hypercholesterolemic adults, Weiße et al. [61] reported that LDL:HDL cholesterol ratio

was beneficially reduced for the subjects consuming lupin proteins (35 g/day). The authors concluded that lupin protein can lower cardiovascular risk, albeit acknowledging that a 35 g/day dose is impracticable [61]. Later, in a double-blind, controlled crossover study performed on 33 hypercholesterolemic human adults, the inclusion of a feasible dose (25 g/day) of *L. angustifolius* seed proteins decreased diastolic blood pressure, pulse at rest, and benefited the LDL:HDL ratio [62].

All in all, bioactive properties of lupin proteins have been proven, and their beneficial effects in subjects with hypercholesterolemia are undeniable. However, the underlying mechanisms remained somewhat elusive, even if some of the above mentioned studies have provided some clues (Fig. 1A). Thus, Weiße et al. [61] and Parolini et al. [56] reported for lupin-protein-fed groups an increase in the expression and activity of SREBP-2, a transcription factor that upregulates the LDL receptors (LDL-R) in hepatocytes. Furthermore, Bettzieche et al. [55] and Parolini et al. [56] reported that expression of *CYP7A1*, a rate-limiting enzyme which converts cholesterol into bile acids, was also up-regulated by the lupin protein diet, suggesting that the cholesterol-lowering effect was also due to an increased excretion of bile salts in feces. On the basis of the hypothesis that specific peptides influence the liver gene expression, Lammi et al. [63] made an important breakthrough in this field by elucidating some of the molecular pathways that play a role in the reported cholesterol-lowering effects. They carried out an *in vitro* assay on a hepG2 cell line, which was treated with two hydrolysates of *L. albus* protein extract. The main finding of the study was that lupin protein hydrolysates can increase the LDL-R concentration and activity, therefore increasing the LDL uptake and reducing cholesterolemia, via the regulation of *SREBP-2* gene expression through the IP3K/Akt/GSK3 $\beta$  metabolic pathway. Another important discovery has been that protein hydrolysates act as statins by decreasing the HMG CoA reductase activity, therefore reducing the *de novo* cholesterol synthesis [63].

In addition to hypocholesterolemic effects, lupin seed proteins are also under current research for their use in preventing or reducing hypertension. Boschini et al. [64,65] tested the ACE (Angiotensin-converting enzyme) inhibiting capacities of the lupin seed conglutins. The ACE enzyme catalyzed the conversion of Angiotensin I to Angiotensin II, the latter being a powerful vasoconstrictor and an inhibitor of the potent vasodilator bradykinin. In this regard, Boschini et al. [64] discovered an ACE-inhibitor activity in lupin proteins, thus offering a powerful alternative to the synthetic ACE inhibitors, which may cause diverse side effects. Among several legume protein-extracts tested, both *L. angustifolius* and soybean extracts had the strongest ACE-inhibiting capacities, and *L. albus* achieved the third position in their rank. Furthermore, using protein isolates, they demonstrated that the  $\beta$ -/ $\alpha$ -conglutin isolate was responsible for that effect. In a subsequent

study, Boschin et al. [65] improved the methodological approach and extended studies to *L. luteus* seed proteins, which were proven to have a higher ACE-inhibiting score than previously tested species. Although the analysis results are promising, there is a lack of understanding whether the active ACE-inhibiting lupine peptides are able to cross the intestine barrier and reach the circulatory system for an activity *in vivo*.

### 3.2. Benefits on diabetes and hyperglycaemia

Diabetes has become one of the most frequent diseases worldwide and the number of cases is expected to reach 400-450 million cases in 2030 [66,67]. The treatment for diabetes is expensive, as can be seen by the total worldwide cost to treat the disease in the year 2000, which ascended to \$132 billion [68]. This pandemic is mainly represented by type-2 diabetes, as it accounts for about 90% of diagnosed cases. A correct nutrition model constitutes the basic treatment and prevention of type-2 diabetes [69]. Leguminous plants are a key factor in the diabetic diet since they provide vegetable protein and dietary fiber, have a low glycemia index, and are rich in bioactive compounds, which can influence glucose metabolism [70]. Furthermore, lupin and soy have been proved to contain specific bioactive peptides, which is an additional advantage when compared to other legumes. To our knowledge, the first breakthrough in lupin protein properties for diabetes prevention is attributable to Magni et al. [43], who managed to prove their hypoglycemic potential through an assay carried out on rats receiving different  $\gamma$ -conglutin concentrations as well as a constant glucose overdose intake. The results showed a significant decrease in postprandial glycemia, which was  $\gamma$ -conglutin dose-dependent. Magni et al. [43] also described the  $\gamma$ -conglutin capacity of binding insulin *in vitro* ( $K_d = 10 \mu\text{M}$ ), by means of an ionic union, which also showed itself to be pH and ionic-strength dependent. Ten years later, Vargas-Guerrero et al. [71] compared isolated  $\gamma$ -conglutin to a common antidiabetic drug (glibenclamide) in streptozotocin-induced diabetic rats. Interestingly,  $\gamma$ -conglutin administered rats had increased insulin levels, higher *Ins-1* expression and reduced blood glucose levels than the control, although these changes were lower than in the glibenclamide group.

A series of studies have also been conducted on humans, although most of them used lupin flour. The first trial testing the hypoglycemic effect of specific lupin proteins was conducted by Bertoglio et al. [72]. In a 4-week controlled-study they demonstrated that isolated *L. albus*  $\gamma$ -conglutin is capable of exerting a long-term hypoglycemic effect in healthy adults when administered prior to a carbohydrate intake, even if no significant variations were observed for the insulin levels. Interestingly, the effects were proven to be exerted with only 50 mg of isolated protein per dose

[72]. The same laboratory hypothesized a pathway explaining hypoglycemic-effects of *L. albus*  $\gamma$ -conglutin after conducting a new *in vitro* experiment, in which the insulin receptor of myocytes was proven to be targeted by the lupin protein [73]. In this study, common intracellular kinases for insulin and  $\gamma$ -conglutin were tested, and the results suggested that  $\gamma$ -conglutin has some insulin-mimetic properties. The way  $\gamma$ -conglutin reaches the blood torrent from the intestine had been somewhat elusive up to then. However, Capraro et al. [74] proposed a putative metabolic pathway for  $\gamma$ -conglutin protein (*L. albus*), and tested it in an *in vitro* and *ex vivo* model of rat intestine. Their results suggested that the protein made its way across the enterocytes via endomembrane vesicles in order to reach the blood torrent circulation. The study had two major outcomes: Caco-2 cells are responsive to  $\gamma$ -conglutin, and the process allowing the protein to reach the blood involves the transcytosis mechanism (Fig. 1B). Capraro et al. [75] have also proved that hepatocytes (G2 cell-line) respond *in vitro* to  $\gamma$ -conglutin (*L. albus*) by a multiple-phosphorylation and internalization of this protein. However, that study sheds no light on the putative effects of  $\gamma$ -conglutin on the glucose metabolism in hepatocytes. A more recent study compared  $\gamma$ -conglutin to the other conglutins in relation to their effect on body and blood parameters of hyperglycemic rats [76]. The authors reported the advantages of conglutins, in particular  $\gamma$ -conglutin, in controlling body weight gain and glycemia.

Finally, it is worth mentioning that some minor seed proteins could also be of interest from a health and welfare point of view. Among them, Bowman-Birk serine protease inhibitors (BBIs) are a good example of this kind of interest as some BBIs have shown anti-carcinogenic or anti-inflammatory effects [39,77]. However, this protease inhibitor activity, thought to be at the origin of health-promoting properties and allowing them to retain biological activity during digestion, also contributes to their anti-nutritional impact and thus reduces seed quality [78]. Likewise, lupin conglutins have also shown some undesirable allergenic properties [79,80]. Lupin-containing food has to be included as an allergen in the ingredient lists of the products sold in the European Union [81] as well as in the United States [82], and it is also a matter of discussion for the Food Standards Australia New Zealand [83,84]. Lupin allergy has a particular cross-reactivity with peanut allergy sufferers. However, even if taking into account the current and potential food uses of lupin proteins, it is obvious that their allergenicity is a relevant issue, this should not dampen interest in their bioactive properties and their use as nutraceuticals. In this context, the available natural resources of the genus *Lupinus* can offer an invaluable source of seed protein variants presenting an enhanced health-promoting activity and/or reduction of undesirable side effects such as allergenicity in the case of conglutins or anti-nutritional properties in the case of BBIs [as shown for pea; 78].

#### 4. Lupin diversity and genomic resources in lupin seed storage proteins

*Lupinus* genus contains several hundred species (200 to 500, according to the authors) distributed in both the Old and the New World [85, and references therein]. However, species are more numerous in the New World, where they spread from Alaska to southern South America, than in the Old World, which contains less than 20 species (Fig. 2). The latter can be divided into two subgroups: (i) the rough-seeded species (mainly found in North-Equatorial Africa and the Mediterranean region), and (ii) the circum-Mediterranean smooth-seeded species. New World species are almost all small-seeded and outcrossing, whereas Old World lupine seeds are larger and mainly self-pollinating [86]. Moreover, Old World species have variable numbers of chromosomes (between 32 and 52) compared to their New World congeners, which have a more stable chromosome content (mainly  $2n=48$ , with, exceptionally,  $2n=96$ , or  $2n=36$ ) [8]. While phylogenetic relationships between species are still not completely resolved, they have been considerably improved during the last fifteen years, providing a much clearer overview of the evolutionary pattern of the genus *Lupinus* (Fig. 2) [85,87,88].

Although lupins are grown in several temperate countries, genomic and transcriptomic data available on lupin species are still scarce, thus limiting the generation of genomic tools for aiding breeding efforts to improve yield and nutritional quality. However, with the generalization of Next Generation Sequencing (NGS) and the growing interests in lupins, the amount and type of available data should significantly increase in the next few years, notably with a progressive release of transcriptomes and the first complete genome of *L. angustifolius* recently sequenced in Australia [89-91]. Concerning genes encoding for conglutins, different approaches at the genomic and transcriptomic scales have been used to identify them. Through the use of cDNA clones libraries construction and sequencing but also using NGS (Roche-454, Illumina, IonTorrent...) and whole genome sequencing, data about gene structure and expression level in lupin seeds have been accumulated during the last 15 years. Today, many sequences are available for different lupin species. The majority of these genes belong to *L. angustifolius* and 16 different genes have been identified: three encode  $\alpha$ -conglutins, seven belong to the  $\beta$ -conglutin family, two are  $\gamma$ -conglutins and four encode  $\delta$ -conglutins [92]. Encouraged by the relatively low-cost of RNA sequencing (RNAseq) and the continuous improvement in sequencing technologies but also in the assembly and analysis tools of transcriptomes, numerous transcriptomes of *L. angustifolius* seeds at different times of development have been produced. Consequently, Foley and her colleagues [92] assumed

that these 16 genes represent the complete family of conglutins in *L. angustifolius*, because none others conglutin known sequences have been identified in the different transcriptomes. Moreover, all the genes identified thanks to the transcriptomic approach were also found in the draft half-genome of *L. angustifolius* [91,93]. This draft provides for the first time the flanking region of the conglutin genes that could be used to study in detail their structure and regulation [91]. The high accessibility of RNAseq, transcriptome assembly and analysis provides an insight into the variability of genes among species but also across the different varieties of one species. Recently, Foley et al. [93] identified conglutin sequences in seed transcriptomes of different lupin species and varieties: *L. angustifolius* (wild variety P27255, domesticated Unicrop and Tanjil varieties), *L. albus* (Andromeda and Kiev mutant), *L. luteus* (Pootalong, a cultivated variety), *L. mutabilis* (variety ID13) and *L. cosentinii* (domesticated variety Erregulla). The results show an identical number of each type of conglutin but the phylogenetic results revealed differences in the conglutin gene evolution. Indeed, relationships between lupine conglutin genes suggest that “ $\alpha$  and  $\gamma$  conglutins diverged prior to lupin speciation, whereas  $\beta$  and  $\delta$  members diverged both prior to and after speciation” [93].

The advent of RNAseq brought a considerable number of advantages for studying gene expression. Although gene expression can be determined using northern blot, qPCR or microarray analyses, RNAseq appeared as a more powerful tool because of its high sensitivity with no limit in terms of genes number. Moreover, RNAseq allows the detection of transcriptional modification (e.g. alternative splicing) and is now common for expression gene studies [94]. Comparison of lupin seed transcriptomes did not reveal any significant variation in expression between the different varieties of *L. angustifolius* and *L. albus* but showed differences between the different species:  $\delta$  conglutins are highly expressed in *L. cosentinii* and *L. luteus* but not in *L. albus*;  $\beta$  conglutins represented more than 35 % of the conglutin transcripts in *L. mutabilis*, *L. angustifolius* and *L. albus* but not in the other species [93]. Furthermore, Foley et al. [93] did not observe a correlation between transcript expression and protein levels of conglutins, suggesting that posttranscriptional control is an important factor in conglutin abundance in lupin species.

In spite of all these studies referring to genomic and transcriptomic data on lupin seeds, it has been extremely difficult to set up a clear database of the conglutins nucleotide and protein sequences because the availability of data remains poor. For example, none of the any conglutin sequences used in the study of Foley et al. [93] are available in Genbank, only the RNAseq reads have been deposited up to date. Moreover, some sequences available on public databases (e.g Genbank, EMBL) and produced by different research groups at different times are identical or very similar. For example,  $\beta$ -conglutin BOYJF7 [95] is completely included in the conglutin BETA7

(AEB33718.1) from Foley et al. [92] with only one mismatch between the two amino acid sequences. In the same way,  $\beta$  conglutin B8Q5G0 [95] is identical to BETA1 conglutin (AEB33712.1; [93]). These findings illustrate the need for a greater availability of data for research teams, but also the necessity for a clear database without any redundancy or ambiguities between sequences.

## 5. Seed proteomics of lupin domestic species

A mass-spectrometry (MS)-based proteomics analysis of *Lupinus* seeds was published by Wait et al. [96], using the 2-DE standard workflow. Proteins were extracted by Tris-HCl buffer (pH 7.5, 4 °C) in order to obtain a total protein extract, and by the “industrial scale” procedure [47] to procure an LPI-E conglutin fraction. They were then separated by 2-DE (pH 4 to 10, and 7.5-17.5% polyacrylamide gradient) and the resolved spots subjected to ESI- tandem-mass spectrometry (MS/MS) analysis. The identification of proteins based on their peptide fragmentation patterns was performed by correlation of uninterpreted MS/MS spectra with entries in SwissProt/TrEMBL using Protein-Lynx Global Server, and when the score was less than 100, additional searches were performed against the NCBIInr database using MASCOT. The percentage of identified proteins by correlation with database sequences was of 65% (47 over 72 spots analyzed), all but two of the positively identified spots corresponding to  $\beta$ -conglutin fragments [96]. Since then, 12 papers have been published, with the differences between them including the species analyzed and the methodology. A few of these studies used label-free quantitative proteomics, but the use of second generation proteomics was scant and third generation approaches have not yet been employed. This would help to go deeper into the proteome coverage.

Up to date, the proteomic analyses of lupin seeds have not considered their different anatomic parts or their physiological status, and were thus exclusively whole or fractionated mature-seed proteomics. Therefore, currently “lupin seed proteomics” means proteomic analysis of the storage tissues/organs, that will not be a handicap to this review as it is focussed on seed storage proteins. There have been few analyses of lupin seed storage proteomes compared to other legumes [97]. A greater agronomic interest in the legume species most studied, and the fact that complete or important genomic data are available, have certainly contributed to that [98]. Indeed, from the beginning it was reported that the correct identification of conglutins depended largely on the availability of new conglutin cDNA sequences and on mature protein ones deduced by taking into account cleavage sites [96,99]. To help with legume proteomics, a legume specific protein database

(LegProt) containing sequences from seven legume species, i.e., *Glycine max*, *Lotus japonicus*, *Medicago sativa*, *Medicago truncatula*, *Lupinus albus*, *Phaseolus vulgaris*, and *Pisum sativum*) was created in 2011 [100]. However, an important number of the available lupin genomic data (see previous section) have not been uploaded into this database, which could be improved in this sense for use in *Lupinus* proteomics.

Seed proteome analysis has also had to face a number of limitations and challenges related to the disparity of seed proteins in terms of quantity, polymorphism and PTM [97]. Thus, storage proteins are extremely abundant in seeds, and this fact could be technically challenging if any analysis is intended for exploring global seed proteome including low abundance proteins. In fact, an overwhelming majority of the peptides identified after proteomic analysis of lupin seeds belonged to the different families of storage proteins as discussed below. But even in the case of an interest restricted to seed storage proteins, quantitative disparity can represent a constraint for the identification of minority isoforms or PTMs, all of these together resulting in broad polypeptide heterogeneity [101]. To address at least partly this constraint, previously described pre-fractionation methods, allowing the separation of different groups of conglutin families in lupins, have also been integrated into proteomics analyses and have proved helpful for validating and complementing direct seed proteome analysis [95,96,99,101,102]. The first proteomic works used 2DE-MS/MS analysis and were performed on *L. albus*, the most common species of lupin consumed by humans, i.e. cultivars Arés [96] and Multitalia [99]. These pioneer works demonstrated the huge potential of these techniques to extend previous knowledge on conglutin peptide diversity and on structure/action relationships. Even if, in agreement with the high number of proteolytic processing sites in  $\beta$ -conglutin, most of the identified spots corresponded to proteolytically cleaved  $\beta$ -conglutin peptides, these studies also improved knowledge on the other conglutins.

In addition to 2DE-gel-based quantitation methods used in these early research works, other MS-based quantitation strategies have been followed for the study of white lupin seed proteomes. The Food Chemistry and Mass Spectrometry (University of Milan, Italy) laboratory performed two liquid chromatography (LC)-MS/MS label-free experiments. The first of them reported a preliminary qualitative and semi-quantitative evaluation on total and fractionated white lupin storage proteins, as well as lupin foods and food ingredients, through conventional HPLC-ESI-MS/MS [102]; and the second one used microfluidic nanoHPLC-Chip Ion Trap-MS/MS for the qualitative characterization and relative quantitation of white lupin conglutins of four *L. albus* cultivars (Adam, Arés, Lucky, and Multitalia) based on ion intensities [103]. This latter approach facilitated an important increase in the reproducibility and sensitivity of protein detection compared to the first one, and provided the characterization of each conglutin group with a similar accuracy to



that obtained from the 2DE-MS/MS analyses done in previous works [96,99]. Furthermore, this internal standard nanoHPLC-Chip-MS/MS method permitted the identification of amino acid substitutions in conglutin sequences and the elaboration of the different expression of  $\beta$ - and  $\gamma$ -conglutin, reporting the existence of cultivar-specific isoforms that are expressed in a differential way both qualitatively and quantitatively.

Besides *Lupinus albus*, early proteomics studies have also considered *L. angustifolius*, another of the main lupin crop species. The first of these studies reported a 2DE sightseen of *L. angustifolius* seed proteome, as well as *L. albus* and *L. luteus* proteomes, but their complete analysis was not the focus of this work and MS/MS analysis was only done for the spots that presented allergenic properties, i.e. IgE reactivity [95]. Most of the identified IgE-binding proteins were  $\beta$ -conglutins (31 spots), but also some  $\gamma$ -conglutins accounted for IgE-reactive spots (5). A later research work on *L. angustifolius* attempted to evaluate the effects of various processing conditions (thermal and mechanical) on seed proteome using 2DE-MS/MS analysis [104]. However, they began by the identification of the main spots separated after 2DE of *L. angustifolius* (cv Boregine) seed proteins, i.e. 57 over 202 spots detected, which were distributed throughout the gel. Both HPLC-Chip-Ion Trap and HPLC-Chip-qTOF, the latter for “*de novo* sequencing” in order to overcome the lack of *L. angustifolius*  $\alpha$ -conglutin sequences at this time, permitted the assignation of spots to the major lupin storage proteins:  $\alpha$ -conglutin (18 spots),  $\beta$ -conglutin (30),  $\gamma$ -conglutin (3) and  $\delta$ -conglutin (4).

Until recently, *L. luteus*, the third domestic lupin species considered for use as a protein source, had not been the subject of proteomic analysis. Ogura et al. [105] performed a systematic study and developed a comprehensive seed protein catalog for *L. luteus* using a combination of 2DE and nano-LC-MS/MS. Taking advantage of EST sequences available for immature seeds of yellow lupin [90], a total of 152 unique proteins were identified and made available in the World-2DPAGE repository (<http://world-2dpagexpasy.org/repository/0066/>). Among them, eight storage proteins which were present as multiple isoforms in the 2DE gels were reported.

The great heterogeneity of conglutins is well reflected in all these proteomic analyses. Thus, a conserved feature of 2DE maps of lupin seed proteins is the frequent presence of rows of spots with a similar molecular weight but varying pI values, which represent different isoforms of the same protein with different degrees of PTMs. Furthermore, these rows of spots are often placed forming clusters of rows of closely spaced MW values, reflecting a differential proteolytic cleavage from common pro-proteins. Among PTMs, N-glycosylation is an important one for conglutin bioactive potential because it greatly influences the structural and functional characteristics of proteins [106,107]. However, PTMs were only briefly outlined in the above-discussed proteomic studies. Recently, an exhaustive analysis of glycosylation modifications of *L. albus*  $\gamma$ -conglutin was

undertaken using high-resolution accurate mass liquid chromatography–multi-stage mass spectrometry [101]. Schiarea and co-authors [101] performed an in-depth molecular characterization of N-glycosylations, analyzing structure, micro-heterogeneity and attachment site of the bound N-glycan(s), and providing extensive coverage of the protein sequence. The glycoproteomic workflow used in this study was based on the preparation of different protease digests, which were then treated or not with N-glycosidase PNGase A or PNGase F, before being analyzed by LC–Orbitrap MS, where MS survey scans were followed by “data-dependent” MS2 or targeted MSn mode. Of the two partially “overlapping” potential N-glycosylation sites in  $\gamma$ -conglutin, Asn<sub>131</sub> and Asn<sub>132→36</sub>, the N-glycan was proven to reside on Asn<sub>131</sub>, which corresponded to the glycosylation site that was predicted using ANTHEPROT software (Table 1; [108]). Among the N-glycans bound to  $\gamma$ -conglutin, the four most abundant variants had both core  $\beta$ 1,2-xylose and core  $\alpha$ 1-3-fucose plant glyco-epitopes, called Cross-reactive Carbohydrate Determinants (CCDs), but corresponding fucose-free variants were also identified as minor components. These CCDs could contribute to the allergenicity of glycoproteins, but the degree of their contribution and their clinical relevance are controversial issues [109]. Schiarea and co-authors [101] also went further in their explanation of the heterogeneity of the  $\gamma$ -conglutin small subunit, identifying, in addition to the C-terminal [110] and the N-terminal [111] peptides described in previous works, two additional N-terminal tryptic peptides differing in the number of serines (one, three or five) at the N-terminus.

The study of the qualitative and quantitative diversity of seed storage proteins at the interspecific level has also been the subject of lupin proteomics, and, in addition to its usefulness for crop improvement through breeding programs, it is an important one for nutraceuticals since different beneficial health effects have been observed between seed proteins from different lupin cultivars (see section 3). A fingerprinting study by Islam et al. [112] using direct Matrix-Assisted Laser Desorption Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry on 25 cultivars of *L. angustifolius* showed considerable variations in conglutins between the cultivars. They demonstrated the feasibility of this technique to accurately perform a high throughput analysis of the proteins of a large number of lupin seed samples. For all 25 cultivars, a total of 364 seed protein mass peaks were clearly identified by MALDI-TOF, including 355 polymorphic protein peaks. The proteomic diversity reported was very high since the number of mass peaks identified for each cultivar varied from 88 to 186, the percentage of total mass peaks present in a few cultivars (three or less) was high (30.5%), and common mass peaks among cultivars only represented 2.4%. This polymorphism is well tuned in with the heterogeneity of polypeptides based on multigenic origin and huge diversity of PTMs in seed storage proteins. The polymorphic protein mass peaks showed themselves to be useful for phylogenetic analysis which was in agreement with DNA marker-based

studies and broadly supported by pedigree information. Based on this phylogenetic relationship, the same research group selected four *L. angustifolius* cultivars (Coromup, Tanjil, Uniharvest and Yorrel) distributed in different groups, and performed a comprehensive proteome analysis of mature seeds using 2D IEF/SDS–PAGE followed by mass spectrometric peptide sequencing [113]. They reported differences in abundance between cultivars for 24.3% of the spots, among which 20.6% were only present in one or two cultivars. By far the greatest variability was observed in  $\beta$ -conglutin since all the identified proteins either present or absent in some cultivars belonged to this group, which were followed in decreasing order by  $\alpha$ -,  $\gamma$ - and  $\delta$ -conglutin. In addition, this research group, using the same methodology as in their 2011 paper [112], analyzed the interaction of genetic and environmental factors on seed protein profiles of five *L. angustifolius* cultivars (Belara, Coromup, Jenabillup, Mandelup and Tanjil), and showed that protein mass peak profile was significantly influenced by cultivars but not by environmental variation and that interactions between both factors were not significant [114].

These proteomic studies of domestic *Lupinus* species set up the basis and clearly established the huge potential of proteomic techniques for screening and studying *Lupinus* germplasm in order to exploit diversity in seed storage proteins with potential beneficial bioactive properties. Information generated by proteomics contributes to an accelerated discovery of potential useful genes/proteins that could play key roles toward the improvement of the presence of bioactive proteins as an important crop trait. Indeed, a major obstacle for crop improvement programs faced by plant breeders is a limited gene/protein pool of domesticated crop species. In this sense, proteomics approaches are proving to be an effective way to speed up crop improvement programs for different traits since proteomic studies reveal the knowledge that is important for understanding the biological functions, have dynamic capabilities, introduce the dimension of post-translational modifications, and contribute to systems biology which further opening up new methods for crop improvement studies via the signaling, regulatory, and metabolic networks underlying plant phenotypes (for review see [115,116]). To date, *Lupinus* germplasm resources remain essentially untapped, and only very few studies have created an analysis scale considering substantial numbers of lupin species and populations, as will be discussed in the next section.

## **6. New directions in lupin seed proteomics research**

Genetic diversity available for lupins should be explored more extensively as it represents a potential source of substantial diversity in seed proteins, as have been described in the case of

soybean or pea for example [78,117,118]. However, as seen above, almost all the canonical proteomic studies on these important bioactive compounds were conducted on only very few varieties from the two Mediterranean lupins, *L. albus* and *L. angustifolius*, which left nearly unexplored both their widely cultivated and naturally available germplasms, and the huge natural lupin diversity adapted to various ecogeographical conditions in the Old and the New World. The few interspecific comparative analyses, previously performed on seed storage proteins, resulted in the same pattern of conglutin families ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ) [32,119,120]. Nevertheless, these analyses also revealed a great diversity in polypeptide electrophoretic patterns of total denatured and reduced conglutins [120-122]. Most Old World lupins exhibited their own pattern, as can be seen in Fig. 3, and interspecific differences were found in all conglutin families, particularly in conglutins  $\alpha$  and  $\beta$ . While some polypeptide components of conglutins  $\alpha$  and  $\gamma$  appeared to be common and highly conserved among the lupin species surveyed, such as cong- $\alpha$ 4 and  $\gamma$ 1, respectively (see Fig. 4), almost all the other conglutin-polypeptide chains with which they are associated were polymorphic among lupins. The latter mostly results from post-translational transformations of polypeptide-precursors [24,32-34,37,120], rather than from a large number of genes, as recently demonstrated by Foley et al. [93] from analyses of conglutins across some lupin species using transcriptomic, protein and comparative genomic approaches. At a lower level, such polypeptide heterogeneity was also observed at the intraspecific level in Mediterranean lupins, as can be seen, for example, among natural populations in the poorly studied *L. micranthus* (Fig. 5).

It is only recently that different canonical proteomic analyses have undertaken to characterize the diversity of lupin seed proteins at the intra- and interspecific level beyond domestic lupin species. In the first of these works, and the only one published in a scientific journal, Islam and co-authors applied direct MALDI-TOF mass spectrometry, a well-established methodology used in some of their previous works [112,114], to determine the protein profile across 33 genotypes spread in 19 lupin species (all species minus the four Mexican ones listed in Fig. 2) [123]. Results revealed an important seed protein diversity with the identification of 630 polymorphic mass peaks. Interestingly, Islam and co-workers [123] found only 23 mass peaks common to more than 10 species, whereas 111 of the 630 mass peaks were specific to 2 or 3 species generally sharing similar phenotypes but also molecular characters. The cultivated species *L. angustifolius*, *L. albus* and *L. luteus* had the highest number of “rare” mass peaks (shared by 2 or 3 species). Moreover, 19 mass peaks were specific to one species and concerned 12 different species. These findings confirm a great diversity throughout the seed protein of the *Lupinus* genus. Diversity of seed protein is also greater in the Old World smooth-seeded lupin compared to the Old World rough-seeded species [123]. Finally, fingerprints obtained by the MALDI-TOF-MS method were used for phylogenetic

analysis and results revealed distinct clusters containing Old World smooth-seeded, Old World rough-seeded and New World species, except for two New World species which were attached to the old World smooth-seeded clade [123]. These results were consistent with the previous ones obtained by other research groups using other phylogenetic criteria (alkaloids, flavonoids, conglutins, internal transcribed spacers, etc.), revealing the value of the MALDI-TOF-MS approach to screen a huge amount of proteins in different samples and to determine phylogenetic relationships between species.

The other communications reporting studies on protein diversity of genomic resources of lupin seeds were presented during the XIV International Lupin Conference held in Milan (Italy) on June 21-26, 2015. Thus, protein diversity of Mexican lupin species (*L. aschenbornii*, *L. campestris*, *L. hintonii* and *L. montanus*) has been studied using SDS-PAGE and 2DE and compared to the proteome of *L. albus* and *L. angustifolius*. According to these authors, these analyses revealed a great heterogeneity between the different conglutin families of lupins with a huge diversity in polypeptides composing conglutins [124]. Similar work has been achieved on the *L. mutabilis* proteome which, when compared to *L. angustifolius*, revealed a high diversity in terms of structure and composition of  $\alpha$  and  $\beta$  conglutins [125]. These great diversities in seed proteins could be explained by post-translational modifications of proteins, which play an important role in the seed protein life cycle [26,123,124].

Regarding this significant interspecific and intraspecific diversity of lupin conglutins found in this limited sample of *Lupinus* (Fig. 2), it is expected that a wider screening of the taxonomic and natural diversity in the genus, using a combination of the more recent approaches and technologies (transcriptomics, proteomics, functional genomics, biological validation processes, etc.) to identify, characterize and investigate the bioactivity of lupin seed storage proteins, will undoubtedly provide novel potential resources beneficial to human health. They will also be useful for selecting and improving current and future edible lupins, as well as for promoting the other conglutins and discovery of other beneficial functions. In addition, a deeper knowledge of the existing *Lupinus* natural germplasm variants for conglutin proteins will assist, along with associated molecular markers, breeding programs.

## 7. Conclusions

The human population is expected to continue growing, so food production needs to match the increase in demand for the forthcoming years. In this context, nutrition studies indicate that legumes are the perfect candidates for contributing to mankind's diet. In parallel, both common belief and research output agree that there is a strong connection between our nutrition habits and our health. This is where lupin seeds stand out among legumes, not only because of their excellent nutritional properties, but for their recently discovered resources as a functional food. Lupin isolated protein extracts can thus be used in food supplements for the prevention and treatment of dysglycemia and hypercholesterolemia.

The use of conventional and canonical proteomic approaches has been essential in unraveling and exploiting lupin seed protein diversity, which goes far beyond gene diversity since the protein level adds to the latter differential proteolytic cleavage of conglutin pro-proteins and a diverse array of glycosylation forms and sites. However, in spite of the promising health promoting effects of lupin seed proteins, the available natural resources of the genus *Lupinus* remain unexplored, thus hampering the discovery of seed protein variants presenting an enhanced health-promoting activity and therefore their use as nutraceuticals. In this direction, upcoming proteomic technologies have been little exploited for lupin seed proteomics. Currently, a wide array of proteomic tools (top-down or bottom-up, gel-based or gel-free, label-based or label-free, PTMs, interactomics,...) allow the large-scale, high-throughput study of the abundance, structure, variability (sequence and PTMs) and function of proteins. Future research should thus aim to determine the lupin species and varieties containing the most bioactive seed proteins, and to characterize the diversity of lupin seed proteins. In this sense, particular seed proteins should be analyzed in relation to their bioactive properties, and the relationships between structure and bioactive properties studied more thoroughly

## **Acknowledgements**

This work was partly supported by the University of Rennes 1 and CNRS (France). J. Keller was supported by a doctoral research grant from the University of Rennes 1 - French Ministry of Higher Education and Research.

## **References**

- [1] Cronk Q, Ojeda I, Pennington RT. Legume comparative genomics: progress in phylogenetics and phylogenomics. *Curr. Opin. Plant Biol.* 2006; 9: 99-103.
- [2] Erbas M, Certel M, Uslu MK. Some chemical properties of white lupin seeds (*Lupinus albus* L.). *Food Chem.* 2005; 89: 341-345.
- [3] Mikić A, Ćupina B, Mihailović V, Krstić D, Antanasović S, Zorić L, Srebrić M. Intercropping white (*Lupinus albus*) and Andean (*Lupinus mutabilis*) lupins with other annual cool season legumes for forage production. *S. Afr. J. Bot.* 2013; 89: 296-300.
- [4] Pastor-Cavada E, Juan R, Pastor JE, Alaiz M, Vioque J. Analytical nutritional characteristics of seed proteins in six wild lupinus species from southern Spain. *Food Chem.* 2009; 117: 466-469.
- [5] Linnemann AR, Dijkstra DS. Toward sustainable production of protein-rich foods: appraisal of eight crops for Western Europe. Part I. Analysis of the primary links of the production chain. *Crit. Rev. Food Sci.* 2002; 42: 377-401.
- [6] Cowling WA, Bevan JB, Tapia ME. *Lupin: Lupinus L.*, Promoting the conservation and use of underutilized and neglected crops. International Plant Genetic Resources Institute, Rome, Italy; 1998.
- [7] Dunn DB. Cytotaxonomy and distribution of the New World lupin species. In: Proceedings of the Third International Lupin Conference, La Rochelle, France; 1984, p.67–85.
- [8] Gladstones JS. Lupins of the Mediterranean region and Africa. Western Australian Department of Agriculture Technical Bulletin 1974; 26: 1-48.
- [9] FAOSTAT (Food and Agriculture Organization of the United Nations, Statistics Division) Database, <http://faostat3.fao.org/>.
- [10] Sujak A, Kotlarz A, Strobel W. Compositional and nutritional evaluation of several lupin seeds. *Food Chem.* 2006; 98: 711-719.
- [11] Duranti M, Consonni A, Magni C, Sessa F, Scarafoni A. The major proteins of lupin seed: characterisation and molecular properties for use as functional and nutraceutical ingredients. *Trends Food Sci. Tech.* 2008; 19: 624-633.
- [12] Khan MK, Karnpanit W, Nasar-Abbas SM, Huma Z, Jayasena V. Phytochemical composition and bioactivities of lupin: a review. *Int. J. Food Sci. Technol.* 2015; 50: 2004-2012.
- [13] Danse B, Roberfroid M, Breslin L, Contor L. Scientific concepts of functional foods in Europe: Consensus document, *Br. J. Nutr.* 1999; 81: S1–S27.
- [14] Duranti M, Morazzoni P. Nutraceutical properties of lupin seed proteins: a great potential still waiting for full exploitation. *Agro Food Ind. Hi Tech* 2011; 22: 20-23.
- [15] Wang W, Mejia D, Gonzalez E. A new frontier in soy bioactive peptides that may prevent age-related chronic diseases. *Compr. Rev. Food Sci. F.* 2005; 4: 63-78.

- [16] Carbonaro M, Maselli P, Nucara A. Structural aspects of legume proteins and nutraceutical properties. *Food Res Int* 2015; 76: 19-30.
- [17] Dijkstra DS, Linnemann AR, van Boekel TA. Towards sustainable production of protein-rich foods: appraisal of eight crops for Western Europe. PART II: Analysis of the technological aspects of the production chain. *Crit Rev Food* 2003; 43: 481-506.
- [18] Osborne TB. *The Vegetable Proteins*. London, UK: Longmans, Green and Co.; 1924.
- [19] Esnault MA, Pichereau V, Klingler J. Variability of the low molecular weight globulin, conglutin  $\delta$ , within lupin species. *Bot. Acta* 1997; 110: 164-171.
- [20] Salmanowicz BP, Przybylska J. Electrophoretic patterns of seed albumins in the Old-World *Lupinus* species (*Fabaceae*): Variation in the 2S albumin class. *Plant Syst. Evol.* 1994; 192: 67-78.
- [21] Duranti M, Restani P, Poniatowska M, Cerletti P. The seed globulins of *Lupinus albus*. *Phytochem.* 1981; 20: 2071-2075.
- [22] Lilley GG, Inglis AS. Amino acid sequence of conglutin delta, a sulfur-rich seed protein of *Lupinus angustifolius* L. Sequence homology with the C-III alpha-amylase inhibitor from wheat. *FEBS Letters* 1986; 195: 235-241.
- [23] Salmanowicz BP, Weder JKP. Primary structure of 2S albumin from seeds of *Lupinus albus*. *Z. Lebensm. Unters. For.* 1997; 204: 129-135.
- [24] Gayler KR, Kolivas S, Macfarlane AJ, Lilley GG, Baldi M, Blagrove RJ, Johnson ED. Biosynthesis, cDNA and amino acid sequences of a precursor of conglutin delta, a sulphur-rich protein from *Lupinus angustifolius*. *Plant Mol. Biol.* 1990; 15: 879-893.
- [25] Blagrove RJ, Gillespie JM. Isolation, purification and characterization of the seed globulins of *Lupinus albus*. *Aust. J. Plant Physiol.* 1975; 2: 13-27.
- [26] Cerletti P, Fumagalli A, Venturin D. Protein composition of seeds of *Lupinus albus*. *J Food Sci.* 1978; 43: 1409-1414.
- [27] Restani P, Duranti M, Cerletti P, Simonetti P. Subunit composition of the seed globulins of *Lupinus albus*. *Phytochem.* 1981; 20: 2077-2083.
- [28] Casero M, Duranti M, Cerletti P. Heterogeneity of subunit composition in lupin globulins. *J. Sci. Food Agr.* 1983; 34: 1113-1116.
- [29] Duranti M, Cerletti P. Subunit composition of proteins from seeds of *Lupinus albus*. *Plant Foods Hum. Nutr.* 1983; 33: 135-138.
- [30] Melo TS, Ferreira RM, Teixeira ARN. The seed storage proteins from *Lupinus albus* seeds. *Phytochem.* 1994; 37: 641-648
- [31] Gayler KR, Wachsmann F, Kolivas S, Johnson ED. Isolation and characterization of protein bodies in *Lupinus angustifolius*. *Plant Physiol.* 1989; 91: 1425-1431.



- [32] Esnault MA, Merceur A, Citharel J. Biosynthèse des protéines de réserve et formation des corps protéiques dans la graine de lupin jaune (*Lupinus luteus* L., Légumineuses). *Agronomie* 1993; 13: 307-316.
- [33] Duranti M, Guerrieri N, Cerletti P, Vecchio G. The legumin precursor from white lupin seed. *Eur. J. Biochem.* 1992; 206: 941-947.
- [34] Duranti M, Horstmann C, Gilroy J, Croy RRD. The molecular basis for N-glycosylation in the 11S globulin (legumin) of lupin seed. *J. Protein Chem.* 1995; 14: 107-110.
- [35] Duranti M, Lovati MR, Dani V, Barbiroli A, Scarafoni A, Castiglioni S, Ponzzone C, Morazzoni P. The  $\alpha'$  subunit from soybean 7S globulin lowers plasma lipids and upregulates liver  $\beta$ -VLDL receptors in rats fed a hypercholesterolemic diet. *J. Nutr.* 2004; 134: 1334-1339.
- [36] Duranti M, Sessa F, Carpen A. Identification, purification and properties of the precursor of conglutin  $\beta$ , the 7S storage globulin of *Lupinus albus* L. seeds. *J. Exp. Bot.* 1992; 43: 1373-1378.
- [37] Duranti M, Gorinstein S, Cerletti P. Rapid separation and detection of concanavalin A reacting glycoproteins: application to storage proteins of a legume seed. *J. Food Biochem.* 1990; 14: 327-330.
- [38] Duranti M, Sessa F, Scarafoni A, Bellini T, Dallochio F. Thermal stabilities of lupin seed conglutin  $\gamma$  protomers and tetramers. *J. Agr. Food Chem.* 2000; 48: 1118-1123.
- [39] Ilgoutz SC, Knittel N, Lin JM, Sterle S, Gayler KR. Transcription of genes for conglutin gamma and a leginsulin-like protein in narrow-leaved lupin. *Plant Mol. Biol.* 1997; 34: 613-627.
- [40] Eaton-Mordas CA, Moore KA. Seed glycoproteins of *Lupinus angustifolius*. *Phytochem.* 1978; 17: 619-621.
- [41] Scarafoni A, Magni C, Duranti M. Molecular nutraceuticals as a mean to investigate the positive effects of legume seed proteins on human health. *Trends Food Sci. Tech.* 2007; 18: 454-463.
- [42] Komatsu S, Hirano H. Plant basic 7 S globulin-like proteins have insulin and insulin-like growth factor binding activity. *FEBS letters* 1991; 294: 210-212.
- [43] Magni C, Sessa F, Accardo E, Vanoni M, Morazzoni P, Scarafoni A, Duranti M. Conglutin  $\gamma$ , a lupin seed protein, binds insulin in vitro and reduces plasma glucose levels of hyperglycemic rats. *J. Nut. Biochem.* 2004; 15: 646-650.
- [44] Czubinski J, Barciszewski J, Gilski M, Szpotkowski K, Debski J, Lampart-Szczapa E, Jaskolski M. Structure of gamma-conglutin: insight into the quaternary structure of 7S basic globulins from legumes. *Acta Crystallogr D Biol Crystallogr.* 2015; 71: 224-238.
- [45] Sironi E, Sessa F, Duranti M. A simple procedure of lupin seed protein fractionation for selective food applications. *Eur. Food Res. Technol.* 2005; 221: 145-150.

- [46] Duranti M, Scarafoni A, Di Cataldo A, Sessa F. Interaction of metal ions with lupin seed conglutin  $\gamma$ . *Phytochem.* 2001; 56: 529-533.
- [47] Wäsche A, Müller K, Knauf U. New processing of lupin protein isolates and functional properties. *Nahrung* 2001; 45: 393-395.
- [48] Heron M P, Hoyert DL, Murphy SL, Xu JQ, Kochanek KD, Tejada-Vera B. Deaths: Final data for 2006. *Natl. Vital Stat. Rep.* 2009; 57, 1-134.
- [49] Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors 2001: systematic analysis of population health data. *Lancet.* 2006; 367: 1747-1757.
- [50] Cam A, González de Mejía E. Role of dietary proteins and peptides in cardiovascular disease. *Mol Nutr Food Res.* 2012; 56: 53-66.
- [51] FDA USA. Food labeling, health claims, soy protein and coronary heart disease. *Fed Regist* 57; 1999, p. 699-733.
- [52] Sirtori CR, Lovati MR, Manzoni C, Castiglioni S, Duranti M, Magni C, Morandi S, D'Agostina A, Arnoldi A. Proteins of white lupin seed, a natural isoflavone-poor legume, reduce cholesterolemia in rats and increase LDL receptor activity in HepG2 cells. *J. Nutr.* 2004; 134: 18-23.
- [53] Marchesi M, Parolini C, Diani E, Rigamonti E, Cornelli L, Arnoldi A, Sirtori CR, Chiesa G. Hypolipidaemic and anti-atherosclerotic effects of lupin proteins in a rabbit model. *Brit. J. Nut.* 2008; 100: 707-710.
- [54] Bettzieche A, Brandsch C, Schmidt M, Weibe K, Eder K, Stangl IG. Differing effect of protein isolates from different cultivars of blue lupin on plasma lipoproteins of hypercholesterolemic rats. *Biosci Biotechnol Biochem.* 2008; 72: 3114–3121.
- [55] Bettzieche A, Brandsch C, Eder K, Stangl, G.I. Lupin protein acts hypocholesterolemic and increases milk fat content in lactating rats by influencing the expression of genes involved in cholesterol homeostasis and triglyceride synthesis. *Mol. Nutr. Food Res.* 2009; 53: 1134-1142.
- [56] Parolini C, Rigamonti E, Marchesi M, Busnelli M, Cinquanta P, Manzini S, Chiesa G. Cholesterol-lowering effect of dietary *Lupinus angustifolius* proteins in adult rats through regulation of genes involved in cholesterol homeostasis. *Food Chem.* 2012; 132: 1475-1479.
- [57] Fontanari GG, Batistuti JP, Cruz RJD, Saldiva PHN, Arêas JAG. Cholesterol-lowering effect of whole lupin (*Lupinus albus*) seed and its protein isolate. *Food Chem.* 2012; 132: 1521-1526.
- [58] Kapravelou G, Martínez R, Andrade AM, Sánchez C, Chaves CL, López-Jurado M, Porres, JM. Health promoting effects of Lupin (*Lupinus albus* var. *multolupa*) protein hydrolyzate and

- insoluble fiber in a diet-induced animal experimental model of hypercholesterolemia. *Food Res.Int.* 2013; 54: 1471-1481.
- [59] Radtke J, Schutkowski A, Brandsch C, Hirche F, Hasenkopf K, Stangl GI. Isolated conglutin  $\gamma$  from lupin, but not phytate, lowers serum cholesterol without influencing vascular lesion development in the apoE-deficient mouse model. *Plant Foods Hum. Nutr.* 2015; 70: 113-118.
- [60] Radtke J, Geissler S, Schutkowski A, Brandsch C, Kluge H, Duranti MM, Stangl GI. Lupin protein isolate versus casein modifies cholesterol excretion and mRNA expression of intestinal sterol transporters in a pig model. *Nutr. Metab.* 2014; 11: 9.
- [61] Weiße K, Brandsch C, Zernsdorf B, Nembongwe GSN, Hofmann K, Eder K, Stangl GI. Lupin protein compared to casein lowers the LDL cholesterol: HDL cholesterol-ratio of hypercholesterolemic adults. *Eur. J. Nutr.* 2010; 49: 65-71.
- [62] Bähr M, Fechner A, Krämer J, Kiehntopf M, Jahreis G. Lupin protein positively affects plasma LDL cholesterol and LDL: HDL cholesterol ratio in hypercholesterolemic adults after four weeks of supplementation: a randomized, controlled crossover study. *Nutr J.* 2013; 12: 107.
- [63] Lammi C., Zanoni C., Scigliuolo GM, D'Amato A, Arnoldi A. Lupin Peptides Lower Low-Density Lipoprotein (LDL) Cholesterol through an Up-regulation of the LDL Receptor/Sterol Regulatory Element Binding Protein 2 (SREBP2) Pathway at HepG2 Cell Line. *J Agr. Food Chem.* 2014; 62: 7151-7159.
- [64] Boschin G, Scigliuolo GM, Resta D, Arnoldi A. ACE-inhibitory activity of enzymatic protein hydrolysates from lupin and other legumes. *Food Chem.* 2014; 145: 34-40.
- [65] Boschin G, Scigliuolo GM, Resta D, Arnoldi A. Optimization of the Enzymatic Hydrolysis of Lupin (*Lupinus*) Proteins for Producing ACE-Inhibitory Peptides. *J Agr. Food Chem.* 2014; 62: 1846-1851.
- [66] Halban PA, Ferrannini E, Nerup J. Diabetes research investment in the European Union. *Nat. Med.* 2006; 12: 70-72.
- [67] Guariguata L. Tracking the global epidemic—new estimates from the IDF diabetes atlas update for 2012. *Diabetes Voice* 2012; 57, 12-15.
- [68] Yach D, Stuckler D, Brownell KD. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nat. Med.* 2006; 12: 62-66.
- [69] Bi Y, Wang T, Xu M. Advanced research on risk factors of type 2 diabetes. *Diabetes Metab Res Rev.* 2012; 28: 32–39.
- [70] Gętek M, Czech N, Muc-Wierzgoń M, Grochowska-Niedworok E, Kokot T, Nowakowska-Zajdel E. The Active Role of Leguminous Plant Components in Type 2 Diabetes. *Evid.-Based Compl. Alt.* 2014; 2014: 293961.

- [71] Vargas-Guerrero B, García-López PM, Martínez-Ayala AL, Domínguez-Rosales JA, Gurrola-Díaz CM. Administration of *Lupinus albus* gamma conglutin (C $\gamma$ ) to n5 STZ rats augmented *Ins-1* gene expression and pancreatic insulin content. *Plant Food Hum. Nutr.* 2014; 69: 241-247.
- [72] Bertoglio JC, Calvo MA, Hancke JL, Burgos RA, Riva A, Morazzoni P, Duranti M. Hypoglycemic effect of lupin seed  $\gamma$ -conglutin in experimental animals and healthy human subjects. *Fitoterapia* 2011; 82: 933-938.
- [73] Terruzzi I, Senesi P, Magni C, Montesano A, Scarafoni A, Luzi L, Duranti M. Insulin-mimetic action of conglutin- $\gamma$ , a lupin seed protein, in mouse myoblasts. *Nutr. Metab. Cardiovasc. Dis.* 2011; 21: 197-205.
- [74] Capraro J, Clemente A, Rubio LA, Magni C, Scarafoni A, Duranti M. Assessment of the lupin seed glucose-lowering protein intestinal absorption by using *in vitro* and *ex vivo* models. *Food Chem* 2011; 125: 1279-1283.
- [75] Capraro J, Magni C, Faoro F, Maffi D, Scarafoni A, Tedeschi G, Duranti M. Internalisation and multiple phosphorylation of  $\gamma$ -Conglutin, the lupin seed glycaemia-lowering protein, in HepG2 cells. *Biochem Biophys Res Commun* 2013; 437: 648-652.
- [76] Capraro J, Magni C, Scarafoni A, Caramanico R, Rossi F, Morlacchini M, Duranti M. Pasta supplemented with isolated lupin protein fractions reduces body weight gain and food intake of rats and decreases plasma glucose concentration upon glucose overload trial. *Food Funct.* 2014; 5: 375-380.
- [77] Clemente A, Arques MC. Bowman-Birk inhibitors from legumes as colorectal chemopreventive agents. *World J. Gastroenterol.* 2014; 20: 10305-10315.
- [78] Clemente A, Arques MC, Dalmais M, Le Signor C, Chinoy C, Olias R, Rayner T, Isaac PG, Lawson DM, Bendahmane A, Domoney C. Eliminating Anti-Nutritional Plant Food Proteins: The case of seed protease inhibitors in pea. *PLoS ONE* 2015; 10: e0134634.
- [79] Guillamon E, Rodriguez J, Burbano C, Muzquiz M, Pedrosa MM, Cabanillas B, Crespo JF, Sancho AI, Mills ENC, Cuadrado C. Characterization of lupin major allergens (*Lupinus albus* L.). *Mol. Nutr. Food Res.* 2010; 54: 1668-1676.
- [80] Jappe U, Vieths S. Lupine, a source of new as well as hidden food allergens. *Mol. Nutr. Food Res.* 2010; 54: 113-126.
- [81] EU (European Union). Commission directive 2007/68/EC, of 27 November 2007, amending Annex IIIa to Directive 2000/13/EC of the European Parliament and of the Council as regards certain food ingredients. *Official Journal of the European Union* L 310; 2007, p.11-14.
- [82] FDA USA. Food and Drug Administration Consumer Health Information. Allergies to a legume called lupin: what you need to know; 2014.

- [83] Smith WB, Gillis D, Kette FE. Lupin: a new hidden food allergen. *Med J Aust*; 2004; 181: 219-220.
- [84] FSANZ (Food Standard Australian New Zealand). Lupin as an Allergen. Administrative Assessment Report 2013; Proposal P1026.
- [85] Ainouche A, Bayer RJ. Phylogenetic relationships in *Lupinus* (Fabaceae: Papilionoideae) based on internal transcribed spacer sequences (ITS) of nuclear ribosomal DNA. *Am. J. Bot.* 1999; 86:590–607.
- [86] Plitmann U. Evolutionary History of the Old World Lupines. *Taxon* 1981; 30: 430-437-
- [87] Drummond CS, Eastwood RJ, Miotto STS, Hughes CE. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): Testing for key innovation with incomplete taxon sampling. *Syst. Biol.* 2012; 61: 443-460.
- [88] Mahé F, Markova D, Pasquet R, Misset MT, Ainouche A. Isolation, phylogeny and evolution of the *SymRK* gene in the legume genus *Lupinus* L. *Mol. Phylogenet. Evol.* 2011; 60: 49–61.
- [89] Kamphuis LG, Hane JK, Nelson MN, Gao L, Atkins CA, Singh KB. Transcriptome sequencing of different narrow-leafed lupin tissue types provides a comprehensive uni-gene assembly and extensive gene-based molecular markers. *Plant Biotechnol. J.* 2015; 13: 14–25.
- [90] Parra-González LB, Aravena-Abarzúa GA, Navarro-Navarro CS, Udall J, Maughan J, Peterson LM, Salvo-Garrido HE, Maureira-Butler IJ. Yellow lupin (*Lupinus luteus* L.) transcriptome sequencing: molecular marker development and comparative studies. *BMC Genomics* 2012; 13: 425.
- [91] Yang H, Tao Y, Zheng Z, Zhang Q, Zhou G, Sweetingham MW, Howieson JG, Li C. Draft Genome Sequence, and a Sequence-Defined Genetic Linkage Map of the Legume Crop Species *Lupinus angustifolius* L. *PLoS ONE* 2013; 8: e64799.
- [92] Foley RC, Gao LL, Spriggs A, Soo LY, Goggin DE, Smith PM, Atkins CA, Singh KB. Identification and characterisation of seed storage protein transcripts from *Lupinus angustifolius*. *BMC Plant Biol.* 2011; 11: 59.
- [93] Foley RC, Jimenez-Lopez JC, Kamphuis LG, Hane JK, Melser S, Singh KB. Analysis of conglutin seed storage proteins across lupin species using transcriptomic, protein and comparative genomic approaches. *BMC Plant Biol.* 2015; 15: 106.
- [94] Frazee AC, Sabunciyan S, Hansen KD, Irizarry RA, Leek JT. Differential expression analysis of RNA-seq data at single-base resolution. *Biostatistics* 2014; 15: 413-426.
- [95] Goggin DE, Mir G, Smith WB, Stuckey M, Smith PMC. Proteomic analysis of lupin seed proteins to identify conglutin beta as an allergen, Lup an 1. *J Agr. Food Chem.* 2008; 56: 6370-6377.

- [96] Wait R, Gianazza E, Brambilla D, Eberini I, Morandi S, Arnoldi A, Sirtori CR. Analysis of *Lupinus albus* storage proteins by two-dimensional electrophoresis and mass spectrometry. *J Agr. Food Chem.* 2005; 53: 4599-4606.
- [97] Miernyk JA. Seed Proteomics. In: Jorrin-Novo JV, Komatsu S, Weckwerth W, Wienkoop S, editors. *Methods in Molecular Biology*, vol. 1072: *Plant Proteomics: Methods and Protocols*. Alemania: Springer; 2013, p. 361-377.
- [98] Champagne A, Boutry M. Proteomics of nonmodel plant species. *Proteomics* 2011; 313: 663-673.
- [99] Magni C, Scarafoni A, Herndl A, Sessa F, Prinsi B, Espen L, Duranti M. Combined 2D electrophoretic approaches for the study of white lupin mature seed storage proteome. *Phytochem.* 2007; 68: 997-1007.
- [100] Lei Z, Dai X, Watson BS, Zhao PX, Sumner LW. A legume specific protein database (LegProt) improves the number of identified peptides, confidence scores and overall protein identification success rates for legume proteomics. *Phytochem.* 2011; 72: 1020-1027.
- [101] Schiarea S, Arnoldi L, Fanelli R, De Combarieu E, Chiabrando C. In-Depth Glycoproteomic characterization of  $\gamma$ -conglutin by high-resolution accurate mass spectrometry. *PLoS ONE* 2013; 8: e73906.
- [102] Locati D, Morandi S, Zanotti M, Arnoldi A. Preliminary approaches for the development of a high performance liquid chromatography/electrospray ionization tandem mass spectrometry method for the detection and label/free semiquantitation of the main storage proteins of *Lupinus albus* in foods. *Rapid Commun. Mass Spectrom.* 2006; 20: 1305-1316.
- [103] Brambilla F, Resta D, Isak I, Zanotti M, Arnoldi A. A label-free internal standard method for the differential analysis of bioactive lupin proteins using nano HPLC-Chip coupled with Ion Trap mass spectrometry, *Proteomics* 2009; 9: 272-286.
- [104] Sirtori CR, Resta D, Brambilla F, Zacherl C, Arnoldi A. The effects of various processing conditions on a protein isolate from *Lupinus angustifolius*. *Food Chem.* 2010; 120: 496-504.
- [105] Ogura T, Ogihara J, Sunairi M, Takeishi H, Aizawa T, Olivos-Trujillo MR, Maureira-Butler IJ, Salvo-Garrido HE. Proteomic characterization of seeds from yellow lupin (*Lupinus luteus* L.). *Proteomics* 2014; 14: 1543-1546.
- [106] Mitra N, Sinha S, Ramya TN, Surolia A. N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends Biochem. Sci.* 2006; 31: 156-163.
- [107] Gomord V, Fitchette AC, Menu-Bouaouiche L, Saint-Jore-Dupas C, Plasson C, Michaud D, Faye L. Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 2010; 8: 564-587.

- [108] Deléage G, Combet C, Blanchet C, Geourjon C. ANTHEPROT: An integrated protein sequence analysis software with client/server capabilities. *Comput. Biol. Med.* 2001; 31: 259-267.
- [109] Altmann F. The role of protein glycosylation in allergy. *Int. Arch. Allergy Immunol.* 2007; 142:99-115.
- [110] Scarafoni A, Di Cataldo A, Vassilevskaia TD, Bekman EP, Rodrigues-Pousada C, Ceciliani F, Duranti M. Cloning, sequencing and expression in the seeds and radicles of two *Lupinus albus* conglutin  $\gamma$  genes. *Biochim. Biophys. Acta* 2001; 1519: 147-151.
- [111] Duranti M, Gius C, Sessa F, Vecchio G. The saccharide chain of lupin seed conglutin gamma is not responsible for the protection of the native protein from degradation by trypsin, but facilitates the refolding of the acid-treated protein to the resistant conformation. *Eur. J. Biochem.* 1995; 230: 886-891.
- [112] Islam S, Ma W, Ma J, Buirchell BJ, Appels R, Yan G. Diversity of seed protein among the Australian narrow-leafed lupin (*Lupinus angustifolius* L.) cultivars. *Crop Pasture Sci.* 2011; 62: 765-775.
- [113] Islam S, Yan G, Appels R, Ma W. Comparative proteome analysis of seed storage and allergenic proteins among four narrow-leafed lupin cultivars. *Food Chem.* 2012; 135: 1230-1238.
- [114] Islam S, Ma W, Ma J, Buirchell BJ, Appels R, Yan G. Genetic and environment interactions of seed storage proteins in narrow-leafed lupine (*Lupinus angustifolius*). *Crop Pasture Sci.* 2012; 63: 1066-1074.
- [115] Salekdeh GH, Komatsu S. Crop proteomics: aim at sustainable agriculture of tomorrow. *Proteomics* 2007; 7: 2976–2996.
- [116] Eldakak M, Milad SIM, Nawar AI, Rohila JS. Proteomics: a biotechnology tool for crop improvement. *Front. Plant Sci.* 2013; 4: 35.
- [117] Natarajan S, Xu C, Bae H, Bailey BA. Proteomic and genomic characterization of Kunitz trypsin inhibitors in wild and cultivated soybean genotypes. *J. Plant Physiol.* 2007; 164: 756-763.
- [118] Natarajan S, Xu C, Bae H, Bailey BA, Cregana P, Capernad TJ, Garrette WM, Luthriaif D. Proteomic and genetic analysis of glycinin subunits of sixteen soybean genotypes. *Plant Physiol. Biochem.* 2007; 45: 436-444.
- [119] Aïnouche A, Guerrieri N, Misset MT, Huon A, Cerletti P. Diversity of seed storage proteins in wild lupin populations from Algeria (North Africa). In: *Proceedings of the International Conference on Plant Proteins from European Crops*, INRA Editions;1998, p.45-49.
- [120] Aïnouche A. Diversité et évolution du genre *Lupinus* L. (Fabaceae). Dissertation Université de Rennes1, France; 1998.

- [121] Przybylska J, Zimniak-Przybylska Z. Electrophoretic patterns of seed globulins in the Old-World *Lupinus* species. *Genet. Resour. Crop Ev.* 1995; 42: 69-75.
- [122] Zimniak-Przybylska Z, Przybylska J. Electrophoretic seed globulin patterns in some New World *Lupinus* species. *Genet. Resour. Crop Ev.* 1997; 44: 57-62.
- [123] Islam S, Ma W, Ma J, Appels R, Yan G. Mass spectrometric fingerprints of seed protein for defining *Lupinus* spp. relationships. *Genet. Resour. Crop Ev.* 2013; 60: 939-952.
- [124] Boschin G, Bermudez TK, Sirtori E, Arnoldi A. Wild Mexican lupins: proteomic analysis of the seeds of four species. In: *Proceedings of the XIV International Lupin Conference*. Milan, Italy; 2015, p.113.
- [125] Aiello G, Boschin G, Arnoldi A. Extensive Characterization of *L. mutabilis* using a combined approach based on 2D-electrophoresis and mass spectrometry. In: *Proceedings of the XIV International Lupin Conference*. Milan, Italy; 2015, p.116.



## FIGURE CAPTIONS

**Fig. 1.** Underlying mechanisms and pathways of hypocholesterolemic and normoglycemic effects of lupin seed proteins. **(A)** Hepatocyte-cholesterol metabolic mechanisms that have been suggested to be altered by lupin protein hydrolysates. Drawn from the results of Parolini et al. [56] and Lammi et al. [63]. **(B)** Suggested pathway for  $\gamma$ -conglutin to exert its normoglycemic effects. Drawn from the results of Capraro et al. [74], Terruzzi et al. [73], Capraro et al. [75] and Vargas-Guerrero et al. [71]. IR= Insulin receptor.

**Fig. 2.** Schematic dendrogram of the *Lupinus* genus. This dendrogram has been redrawn and adapted using data from Drummond et al. [87]. Each branch corresponds to a different *Lupinus* species. The names indicated here correspond to the *Lupinus* species used in canonical proteomic approaches, discussed in this review. Dashed lines indicate predicted position of *Lupinus* species that were not reported in Drummond et al. [87]. General geographic regions are also indicated. Flo: Floridian; Circum-Med: Smooth-seeded Circum-Mediterranean; Afri-Med: Rough-seeded North Equatorial Africa/Mediterranean.

**Figure 3.** Diversity of seed storage proteins amongst 5 different Old World lupin species. Electrophoretic patterns of total conglutins under denatured and reduced conditions (R-SDS-PAGE) isolated from mature seed samples of 13 populations and varieties, originating from different localities in North Africa and eastern Mediterranean [120]. 1= *L. pilosus* (Pi); 2= *L. albus* (Al); 3, 4, 5, 6 and 7 = *L. micranthus* (Mic); 8 and 9= *L. luteus* (Lu); 10, 11, 12 and 13= *L. angustifolius* (Ang).

**Figure 4.** Interspecific variability of mature seed storage proteins within conglutin families among the Old World lupins. Comparative diagrams (on the right) summarizing the polypeptide chain components of  $\gamma$ -,  $\alpha$ - and  $\beta$ -conglutins identified in 5 different species (one sample per species), resulting from one- and two-dimensional polyacrylamide gel electrophoreses (PAGE) analyses, under denaturing conditions of non-reduced (SDS-PAGE) and reduced (R-SDS-PAGE) of the protein samples [120]. The non-reduced subunit (S.U) structure is indicated in boxes on the left for each conglutin family. Mi = *L. micranthus*; An = *L. angustifolius*; Lu = *L. luteus*; Al = *L. albus*; Pi = *L. pilosus*.

**Figure 5.** Within species variability of mature seed storage proteins in the Mediterranean lupin, *L. micranthus*. A comparative diagram summarizing the polypeptide chain components of  $\alpha$ - and  $\beta$ -

conglutins identified in 5 natural populations originating from different localities in North Africa. Characterization of subunits and polypeptide chains resulting from different steps of conglutin isolation, differential fractionation, and one- and two-dimensional polyacrylamide gel electrophoreses (PAGE) under denaturing conditions of non-reduced (SDS-PAGE) and reduced (R-SDS-PAGE) protein samples [119,120]. Seed samples 3, 4, 5, 6 and 7 originated from Berrouaghia, L'Arbatache, Beloua, Bainem and Boutlelis, respectively, in Algeria.

Accepted manuscript

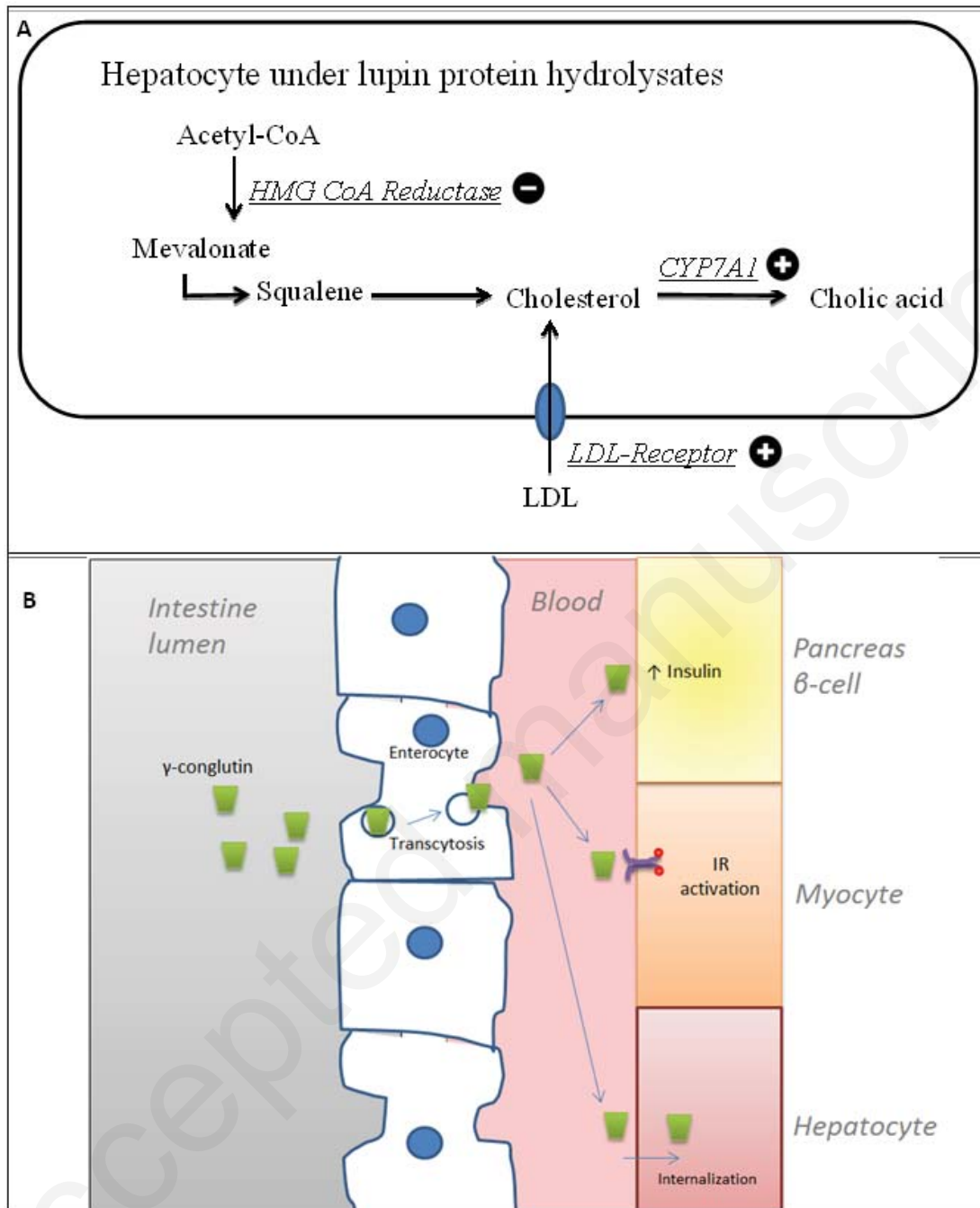


Fig. 1

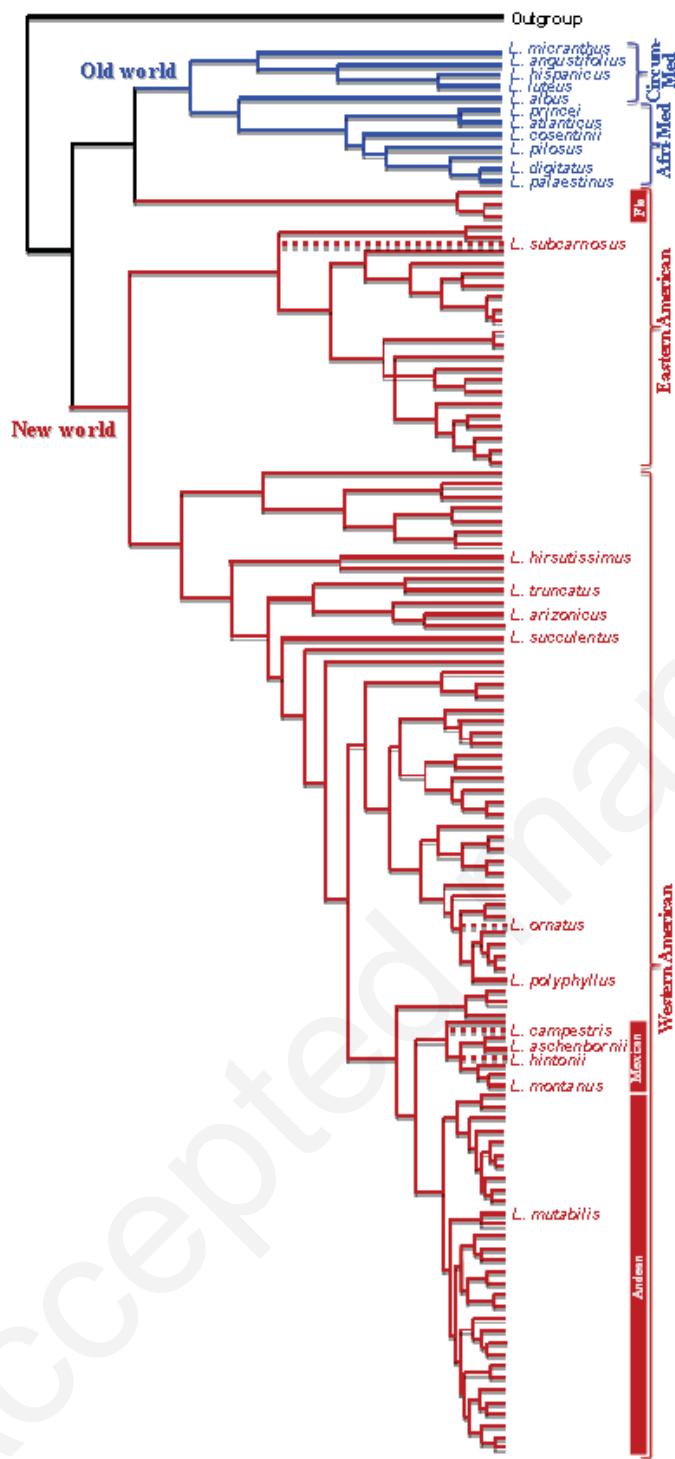


Fig. 2

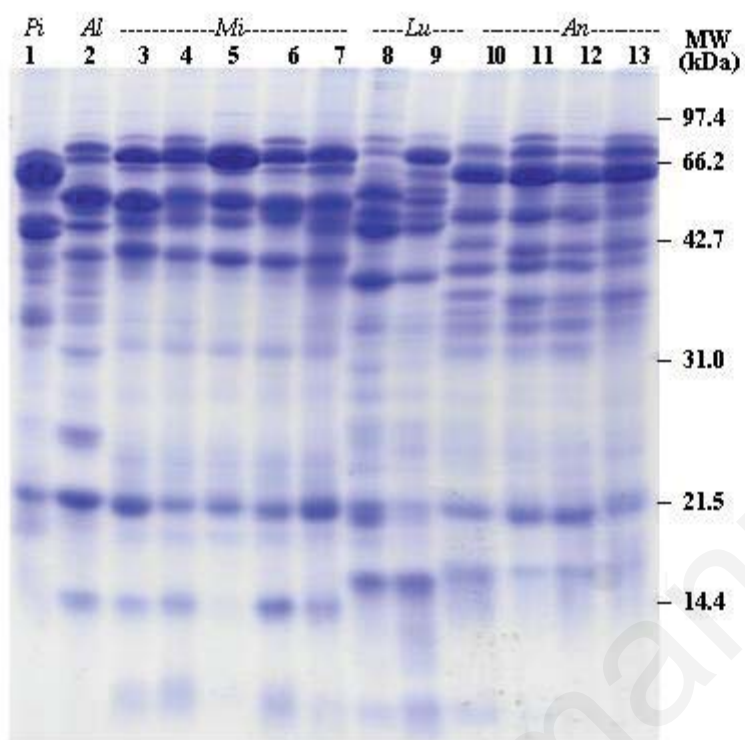


Fig. 3

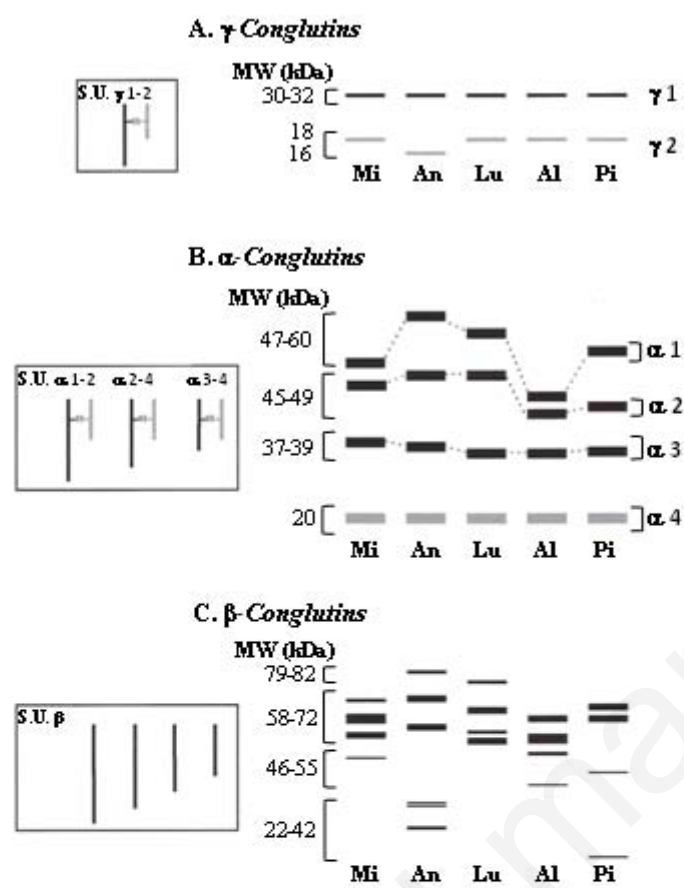


Fig. 4

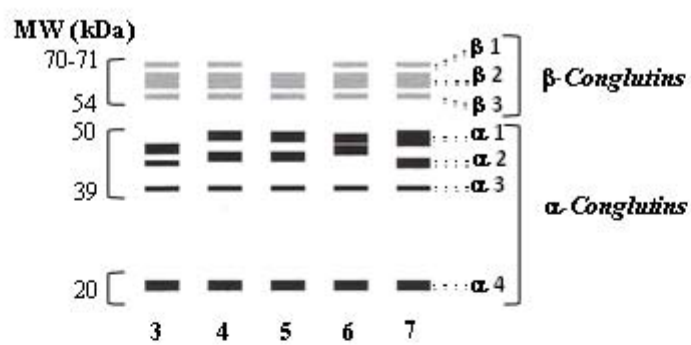


Fig. 5

**Table 1. Key features of lupin conglutin genes and proteins**

Group (Family)	Protein primary structure	Member	Nucleotide sequence <sup>a</sup>		Native pro-polypeptide <sup>a</sup>					Monomer composition <sup>b</sup>					
			GenBank	Size (bp)	UniProtKB	Size (aa)	MW (kDa)	Signal peptide <sup>c</sup>	pI <sup>d</sup>	Gly <sup>e</sup>	Subunit	MW (kDa)	pI	S-S <sup>f</sup>	Gly <sup>g</sup>
α-conglutinin (11S or legumin)	Hexamer	ALPHA 1	HQ670 406	17 67	F5B8V 6	50 6	57.8	Y (19 aa)	5.1	397 (NGSA) 439 (NFSY)	Acidic	42 -5 2	4.5- 4.7	Y	Y / N
		ALPHA 2	HQ670 407	21 43	F5B8V 7	64 3	74.2	Y (22 aa)	4.9	No	Basic	20 -2 2	6.7- 8.6	Y	No
		ALPHA 3	HQ670 408	19 70	F5B8V 8	58 5	67.4	Y (23 aa)	5.1	No	Basic	20 -2 2	6.7- 8.6	Y	No
β-conglutinin (7S or vicilin)	Trimer	BETA 1	HQ670 409	19 90	F5B8V 9	61 1	71.9	Y (30 aa)	5.7	434 (NISL) 519 (NASS)	Large	53 -6 4	5.1- 5.7	No	Y
		BETA 2	HQ670 410	19 49	F5B8 W0	60 3	70.7	Y (30 aa)	5.5	504 (NASS)					
		BETA 3	HQ670 411	18 96	F5B8 W1	58 0	68.2	No	5.6	229 (NGSA) 488 (NASS)	Medium	25 -4 6	5.3- 8.4	No	Y / N
		BETA 4	HQ670 412	19 39	F5B8 W2	59 0	69.4	No	5.9	498 (NASS)					
		BETA 5	HQ670 413	20 80	F5B8 W3	63 7	75.2	Y (30 aa)	5.6	544 (NASS)					
		BETA 6	HQ670 414	19 49	F5B8 W4	59 3	70.1	No	6.1	242 (NGSA) 500 (NASS)	Small	17 -2 0	4.2- 5.0	No	Y / N
		BETA 7	HQ670 415	19 81	F5B8 W5	60 5	71.6	Y (30 aa)	5.5	406 (NESI) 513 (NASS)					



γ-conglutinin (7S)	Tetramer	GAM MA 1	HQ670 416	15 28	Q4236 9	44 9	48.9	Y (27 aa)	7.5	130 (NNTC)	Large	29	8.2–8.9	Y	Y
		GAM MA 2	HQ670 417	15 65	F5B8 W7	43 1	47.2	No	6.4	116 (NNTC)	Small	17	5.8–6.6	Y	No
δ-conglutinin (2S)	Monomer	DELT A 1	HQ670 418	77 2	F5B8 W8	15 2	17.8	Y (22 aa)	5.3	No	Large	9	4.1–4.3	Y	No
		DELT A 2	HQ670 419	69 1	Q9923 5	15 3	17.8	Y (22 aa)	5.1	No					
		DELT A 3	HQ670 420	71 2	F5B8X 0	14 9	17.4	Y (22 aa)	6.0	No	Small	4	n.d.	Y	No
		DELT A 4	HQ670 421	58 4	F5B8X 1	96	10.7	Y (22 aa)	8.2	No					

Data from <sup>a</sup> *Lupinus angustifolius* or <sup>b</sup> *Lupinus albus* [11] conglutins.

<sup>c</sup> As reported in the UniProtKB database (<http://www.uniprot.org/uniprot/>).

<sup>d</sup> Calculated using ANTHEPROT software (<http://antheprot-pbil.ibcp.fr>) [105].

<sup>e</sup> Amino acid position and sequences of potential N-glycosylation sites (N-{P}-[ST]-{P}) determined using ANTHEPROT software.

<sup>f</sup> Covalently linked by disulphide (S-S) bonds.

<sup>g</sup> Glycosylation. Y/N: not all the polypeptides are glycosylated.

S: Sedimentation coefficient. Y: Yes. aa: Amino acids. bp: Base pairs. n.d.: Not determined.

## Significance

Some important topics concerning storage proteins from lupin seeds are presented and analyzed in an integrated way in this review. Proteomic approaches have been essential in characterizing lupin seed protein diversity, which goes far beyond gene diversity since the protein level adds to the latter differential proteolytic cleavage of conglutin pro-proteins and a diverse array of glycosylation forms and sites. Proteomics has also proved helpful for screening and studying *Lupinus* germplasm with the future aim of exploiting and improving food production, quality, and nutritional values.

Accepted manuscript



Graphical abstract

Accepted manuscript

## Highlights

- Lupin seed proteins stand out among legumes for their bioactivity potential.
- Proteomics has been essential in unraveling lupin conglutin diversity.
- Conglutin diversity is mainly due to differential proteolysis and glycosylation.
- *Lupinus* germplasm remains basically unexploited for seed protein diversity.

Accepted manuscript