the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) doi: 10.1111/tpj.16329

RESOURCE

A comprehensive metabolomics and lipidomics atlas for the legumes common bean, chickpea, lentil and lupin

Mustafa Bulut[1], Regina Wendenburg[1], Elena Bitocchi[2], Elisa Bellucci[2], Magdalena Kroc[3], Tania Gioia[4], Karolina Susek[3], Roberto Papa[2], Alisdair R. Fernie[1,5,*] and Saleh Alseekh[1,5,*]

[1]*Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany,*
[2]*Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, via Brecce Bianche, Ancona 60131, Italy,*
[3]*Legume Genomics Team, Institute of Plant Genetics, Polish Academy of Sciences, Strzeszynska 34, Poznan 60-479, Poland,*
[4]*School of Agricultural, Forestry, Food and Environmental Sciences, University of Basilicata, Potenza 85100, Italy, and*
[5]*Center for Plant Systems Biology and Biotechnology, 4000, Plovdiv, Bulgaria*

Received 23 December 2022; revised 9 May 2023; accepted 24 May 2023.
*For correspondence (e-mail alseekh@mpimp-golm.mpg.de; fernie@mpimp-golm.mpg.de).

SUMMARY

Legumes represent an important component of human and livestock diets; they are rich in macro- and micronutrients such as proteins, dietary fibers and polyunsaturated fatty acids. Whilst several health-promoting and anti-nutritional properties have been associated with grain content, in-depth metabolomics characterization of major legume species remains elusive. In this article, we used both gas chromatography–mass spectrometry (GC-MS) and liquid chromatography–mass spectrometry (LC-MS) to assess the metabolic diversity in the five legume species commonly grown in Europe, including common bean (*Phaseolus vulgaris*), chickpea (*Cicer arietinum*), lentil (*Lens culinaris*), white lupin (*Lupinus albus*) and pearl lupin (*Lupinus mutabilis*), at the tissue level. We were able to detect and quantify over 3400 metabolites covering major nutritional and anti-nutritional compounds. Specifically, the metabolomics atlas includes 224 derivatized metabolites, 2283 specialized metabolites and 923 lipids. The data generated here will serve the community as a basis for future integration to metabolomics-assisted crop breeding and facilitate metabolite-based genome-wide association studies to dissect the genetic and biochemical bases of metabolism in legume species.

Keywords: legumes, chickpea, common bean, lentil, lupin, metabolomics, lipidomics.

INTRODUCTION

Legumes are one of the major sources of carbohydrates and proteins for human consumption. The necessity to combat climate change and human health concerns led to the promotion of the food transition towards plant-based diets. Recently, this generated a strong interest in breeding and the conservation of legume genetic resources. Additionally, pulses serve as vitamin and mineral sources, as well as exerting health-beneficial effects that originate from various specialized metabolites (Martin, 2018; Rebello et al., 2014). While the main purpose of grains is human consumption, the by-products such as straw and leaves can serve as protein- and nutrient-rich fodder in feed applications (Capper, 1990). Given this fact, determining the chemical composition and their underlying genetic basis can pave the way to improve legume crop species for both human and livestock diets.

Crop domestication tends to lead to an immense reduction in the allelic diversity of gene pools in modern cultivars (Fernie et al., 2006). In essence, crop genetic resources can provide fundamental sources for agriculture and boost innovation in the food sector. During the last decades, metabolomics techniques became a powerful tool in assessing metabolite levels in a wide range of populations among all kingdoms of life (Bar et al., 2020; Hall et al., 2002; Sebastiana et al., 2021), provided necessary precautions are taken (Alseekh, Aharoni, et al., 2021). The approach has been extensively defined (Alseekh & Fernie, 2018; Fiehn, 2002; Saito & Matsuda, 2010) and used to unravel various metabolic fingerprints among plant

species, e.g., tobacco (*Nicotiana benthamiana*) (Drapal et al., 2022), tomato (*Solanum lycopersicum*) (Tohge & Fernie, 2015; Zhu et al., 2018), rice (*Oryza sativa*) (Gong et al., 2013) and maize (*Zea mays*) (Wen et al., 2014). In the field of food science, metabolomics is used to determine food quality and authenticity (Cubero-Leon et al., 2014). The generation of metabolic compositional resources is essential for breeding programs necessary for improving the quality of future crops. Several resources have been published in recent years for important crop species, including a metabolite database for root, tuber and banana (*Musa spp.*) crops (Price et al., 2020), maize (Desmet et al., 2021) and rice (reviewed in Hong et al. (2019)). To date, metabolic compositional resources are missing for pulses; therefore, metabolic annotations are emphasized here and accumulation sites for legume crop improvement are elucidated tissue-wise.

The Leguminosae family produces a plethora of structurally and functionally diverse metabolites, including chemical compound classes of alkaloids, non-protein amino acids (NPAAs), terpenoids, polyphenols, polyacetylenes and fatty acids, as well as polyketides and carbohydrates (Wink, 2008, 2013). The vast majority of these compounds are classified as specialized metabolites and synthesized in response to biotic and abiotic stresses (He et al., 2022; Heiling et al., 2010; Schulz et al., 2016; Wink, 1983). In addition, those present in medicinal plants, likely belonging to species with the richest metabolite diversity, are known to have a broad range of pharmacological and toxicological properties (Brown Jr & Trigo, 1995; Wink, 1988, 1993; Wink & Schimmer, 2010). Historically, prior to the emergence of molecular systematics, chemical profiles were used as a systemic framework to reflect phylogeny, a concept called chemotaxonomy, as numbers of specialized metabolites are restricted to subfamilies, tribes and genera. Among these, the subfamily Faboideae in the Leguminosae family is composed of several clades with distinct metabolic profiles. Phenolics such as isoflavones are unique to this subfamily, while others, like flavones and anthocyanins, are widely distributed. Contrarily, coumarins are more patch-wise distributed, with a high frequency in the inverted repeat-lacking (IRL) clade and some Phaseoleae. On the other hand, terpenoids, such as saponins, steroids and carotenoids, are distributed across the Leguminosae family. Lastly, alkaloid classes are divergently distributed across clades, e.g., quinolizidine alkaloids (QAs), which occur almost in all taxa in the genistoid clade and other basal taxa in Faboideae.

In this study, we assessed the metabolic diversity in five grain legume species commonly growing in Europe, namely, *Phaseolus vulgaris*, *Cicer arietinum*, *Lens culinaris*, *Lupinus albus* and *Lupinus mutabilis*, in the intelligent collection recently developed for these five species (Cortinovis et al., 2021; Guerra-García et al., 2021; Kroc et al., 2021; Rocchetti et al., 2022). Overall, these species represent a cross-section in terms of their potential value for sustainable food production and are linked to European food traditions and needs. At a molecular level, they show extremes regarding their genome sizes and comprehensive panoplies of genomic resources (Bellucci et al., 2021). In addition, the selected species reflect the divergence of major clades, e.g., the 50-kb inversion clade, the NPAA-accumulating (NPAAA) clade and the IRL clade of the Faboideae subfamily, to provide a broad range of genetic as well as biochemical diversity across the Faboideae. Using both GC-MS and LC-MS we were able to annotate over 3400 metabolites covering the major chemical classes, including polyphenols, alkaloids and terpenoids, and observed species- and tissue-specific abundances. The knowledge of these metabolites on different levels can be implemented for different agronomical and dietary application purposes such as plant resilience, fodder for livestock and human consumption, respectively. In addition to analyzing their metabolome we used the data to reconstruct the major enzymatic steps involved in their biosynthetic pathways. Finally, the metabolic profiles described here serve as a resource for complementary characterization of the genetic diversity including quantitative genetic approaches and breeding. Their use as a baseline for future work on broad genetic resources will pave the way for the full elucidation of their biosynthetic pathways and of the regulatory mechanisms which underpin them. This will ultimately provide tools by which we can improve grain legume crops.

RESULTS AND DISCUSSION

Overall metabolome diversity in legume species

In order to explore the biochemical diversity of major legume species, three tissues (root, leaf and seed) from 22 genotypes, namely, four *L. albus*, three *L. mutabilis*, five *L. culinaris*, five *C. arietinum* and five *P. vulgaris* genotypes, were harvested and subjected to comprehensive metabolite analysis using liquid and gas chromatography coupled to high-resolution mass spectrometry (Figure 1). Metabolite profiling revealed a large metabolic diversity within and between the investigated legume species across different tissues (Figure 1, Table S1). The targeted and non-targeted metabolomic datasets obtained from polar GC-MS, semi-polar LC-MS/MS and non-polar LC-MS allowed the detection of a total of 224, 2283 and 923 metabolites, respectively. A total of 902 metabolites were identified by matching to an in house library for the comparison of their chromatographic and fragmental behaviors to those of the commercial standards (see Experimental Procedures). Between 12 and 26% of the identified metabolites were detected in all species and tissues, while the majority of metabolites (mainly specialized metabolites)

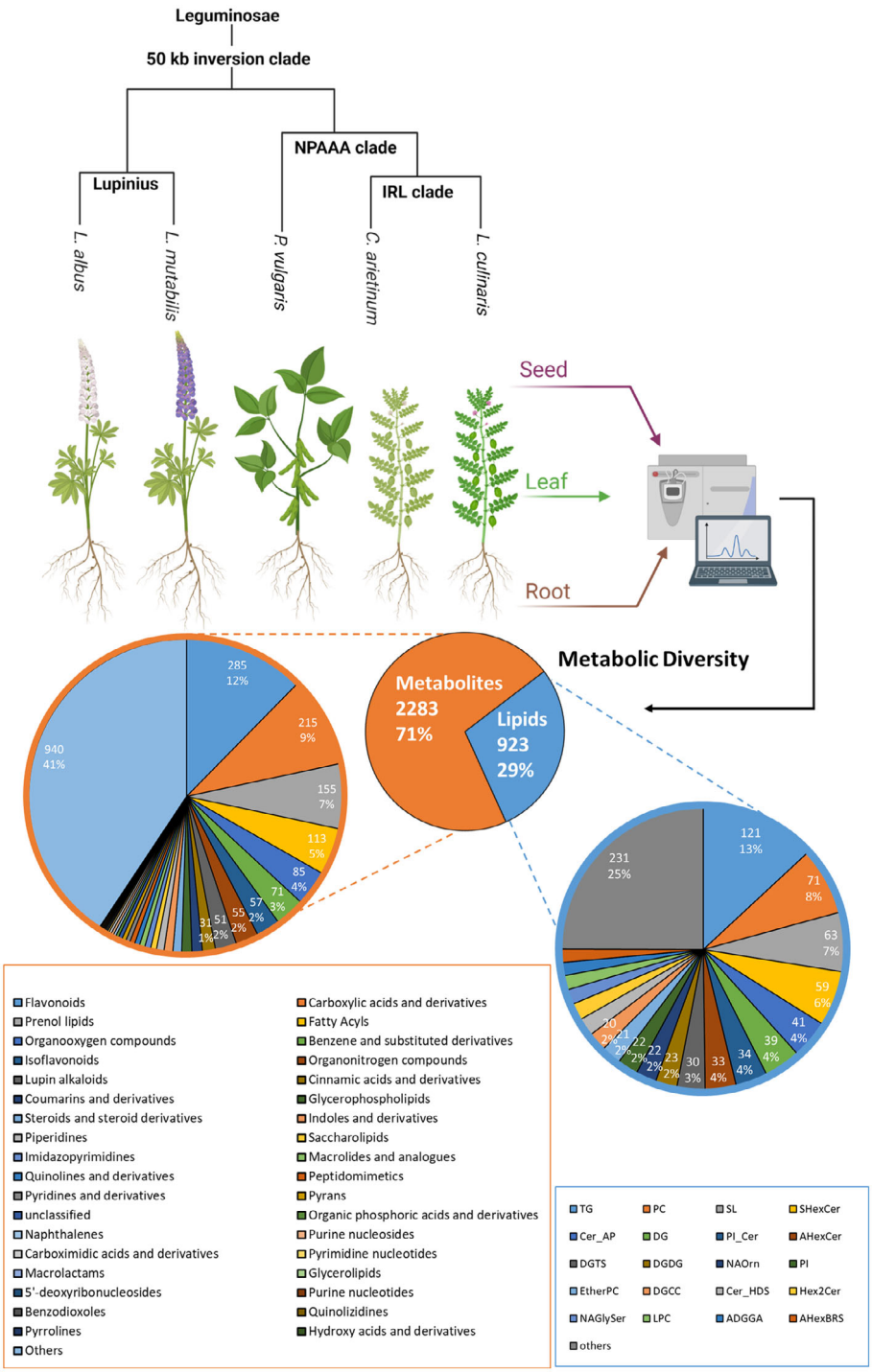


Figure 1. Experimental overview. The scheme illustrates the diverse legume species selected in this study. Several accessions of each species depicted were analyzed for their metabolic composition tissue-wise across several analytical platforms, resulting in 3430 annotated metabolites, further separated into lipidomic (n_l = 923), metabolomic (n_m = 2283) and derivatized (n_d = 224, not depicted here) annotated compounds grouped based on their chemical classes. TG = triacylglycerol, PI = phosphatidylinositol, PC = phosphatidylcholine, DGDG = digalactysoldiacylglercol, DG = diacylglycerol, SL = saccharolipid, SHexCer = sulfatide, Cer_AP = hexosylceramide alpha-hydroxy fatty acid-phytospingosine, PI_Cer = ceramide phosphoinositol, AHexCer = acylhexosylceramide, DGTS = diacylglyceryl trimethylhomoserine, NAOrn = *N*-acyl ornithine, EtherPC = ether-linked phosphatidylcholine, DGCC = diacylglyceryl-3-*O*-carboxyhydroxymethylcholine, Cer_HDS = ceramide hydroxy fatty acid-dihydrosphingosine, Hex2Cer = dihexosylceramide, NAGlySer = *N*-acylglycylserine, LPC = lysophophatidylcholine, ADGGA = acyl diacylglyceryl glucuronide, AHexBRS = acylhexosyl brassicasterol.

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2023), doi: 10.1111/tpj.16329

were detected in a tissue/species-specific manner, such as certain soyasaponin B glycosides (mainly detected in seeds), hydroxylupanine (specific to lupin species) and 3′-*O*-methylmyricetin glycoside (in chickpea seeds). The annotated metabolites showed considerable metabolic diversity and can largely be classified into multiple compound classes, including lipids, sugars, amino acids, organic acids and specialized metabolites such as sparteine, lupanine, alkaloids, isoflavonoid *C*/*O*-glycosides and triterpene saponins (Tables S1, S2 and S3).

Principal component analysis (PCA) was performed on the metabolite datasets in order to broadly explore the metabolic variation of different legume species in different tissues. Results showed that PC1 and PC2 explain 22.97 and 10.44% of the metabolic variation, respectively (Figure 2). The major discriminate in sample separation, as depicted in PC1, originates from variation at the tissue level by separation of seeds from root and leaves, whereas PC2 strongly separates leaves and roots (Figure 2). Overall, the observed intratissue variation resembles the phylogenic distance, with the lupin species being clearly separable from other legume species. Analysis of the loadings for PC2 indicates that the tissue-specific accumulation of compounds classified as triterpene saponins, such as soyasapongenol E and B derivatives, have the lowest PC scores, thus highlighting their importance in these tissues, whereas some lipids, such as digalactosyldiacylglycerols (DGDGs), and flavonoid glycosides have positive scores and are highly abundant in leaf tissue. Further, the observed species-wise separation is underlined by species-specific specialized metabolites, such as lupin alkaloids. By contrast, the loadings of PC1 led to clear separation of seeds from other tissues, with a major contribution of lipids, such as triacylglycerols (TGs; Tables S4 and S5).

Metabolic diversity in primary metabolites and their derivatives across legumes

Sugars, amino acids, organic acids and lipids represent major calorie sources for human consumption and have important applications in fodder production. In addition, they have essential roles in determining nutritional quality.

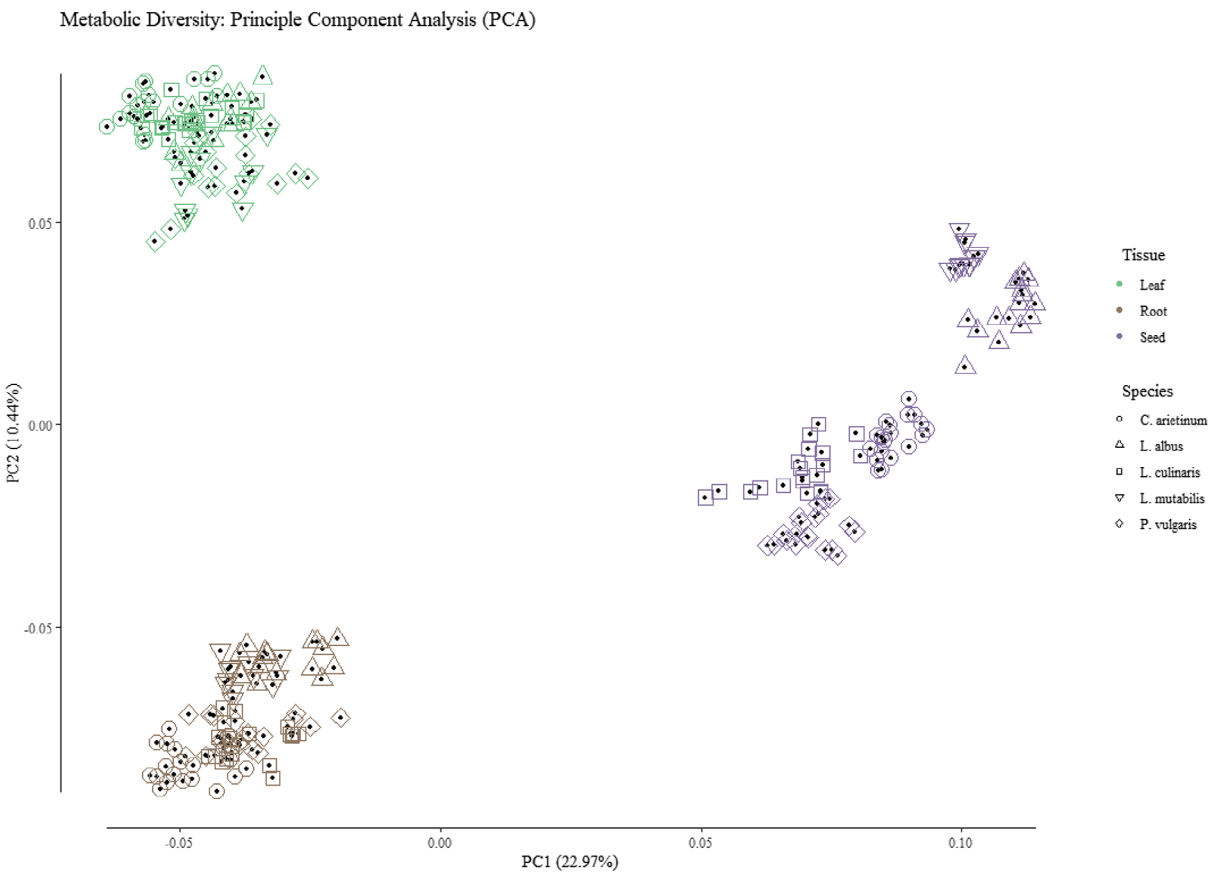


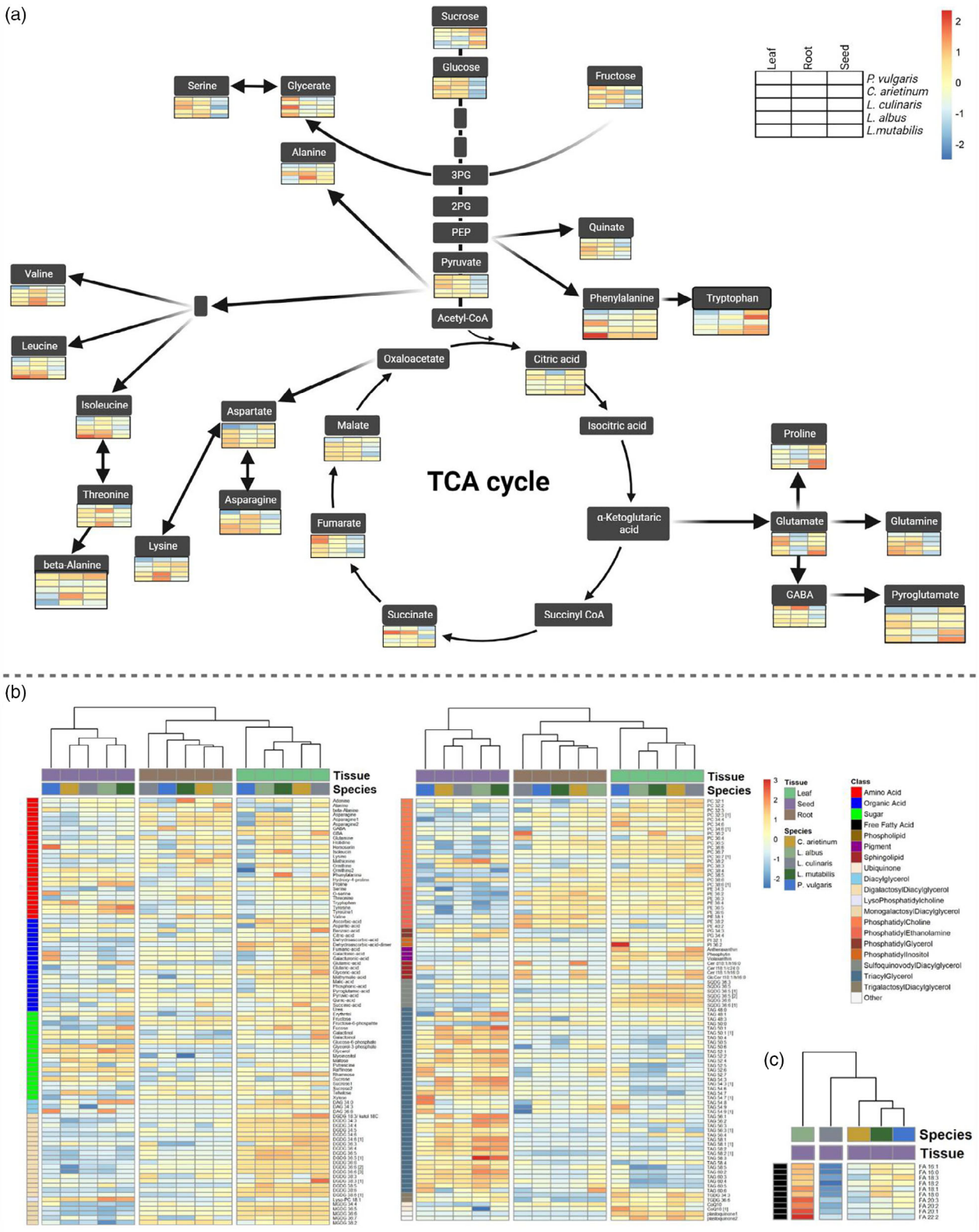
Figure 2. Multivariate analysis of global metabolic changes. PC1 and PC2 of the principal component analysis (PCA) including all detected metabolites across the analytical platforms are depicted over the *x*-axis and the *y*-axis, respectively. Of the total variation, 33.41% is explained by the score plot (PC1 = 22.97% and PC2 = 10.44%). Dots represent individual sample extracts. Colors and shapes indicate tissue and species, respectively. Green = leaf, brown = root, purple = seed, circle = *C. arietinum*, triangle = *L. albus*, square = *L. culinaris*, flipped triangle = *L. mutabilis*, diamond = *P. vulgaris*.

Compared to other grain crops, e.g., cereals, which are low in essential amino acids like lysine, legumes contain appropriate levels of lysine and ensure enhanced protein content in cereal-based diets (Iqbal et al., 2006). In order to provide an overview on the abundance of essential compounds that are involved in plant growth and development as well as important quality components, we detected and quantified using the GC-MS approach 20 organic acids, 26 amino acids, 19 sugars and several other derivatized compounds (Figure 3, Table S2). Overall, results indicate that the highest levels of organic acids such as benzoic acid, malic acid and glutaric acid are found in leaves (Figure 3, Table S4). In *C. arietinum*, tricarboxylic acids such as succinate showed the highest abundance in both leaves and roots, but low levels in seeds. Other important tricarboxylic acids such as fumarate were found to be highly accumulated in *P. vulgaris* and *C. arietinum* leaves. Unspecific distribution of amino acids across species and tissues was observed; for instance, lysine, which is known as the precursor of QAs (reviewed in Boschin and Resta (2013)), showed high accumulation in *L. albus* and *L. mutabilis*. However, we noticed low levels of lysine in seeds, which coincide with the low lysine decarboxylase (LDC) rates in seeds and the observation that the major seed QA constituents are translocated to the seeds (Czepiel et al., 2021; Otterbach et al., 2019). Further, the aromatic amino acids phenylalanine and tryptophan display contrasting accumulation in seeds and leaves across different legume species, indicating potential different chorismate preferences towards the synthesis of either amino acid through arogenate. These amino acids are of particular interest since both are major precursors of downstream specialized metabolite diversification (Tohge et al., 2013). Given the fact that phenylalanine is a major contributor to fuel the phenylpropanoid pathway, its role in driving production towards isoflavonoid production is crucial. At the tissue level, data showed that the amino acids alanine, valine and threonine were mostly accumulated in roots, while proline and tryptophan were mainly accumulated in seeds. In the case of alanine, accumulation has been reported in several plant tissues; for example, in *Hordeum vulgare* it accumulates to higher levels under anaerobic conditions (Good & Muench, 1993), while the aspartate concentration decreases, indicating transamination of pyruvate to alanine as well as catabolic reduction of aspartate to succinate. Sugars are major sources of calories; our results showed that the majority of sugars, such as maltose, raffinose, rhamnose and trehalose, accumulated to the highest levels in seeds. By contrast, fructose, and fructose-/glucose-6-phosphate accumulated in leaves. In legumes, some of the sugars, such as raffinose and fructose, cannot be fully digested and absorbed, leading to the undesired production of gasses by fermentation (Bornet, 1994; Zartl et al., 2018), whereas others are more desired, such as galactose and arabinose (Buxton et al., 1987). As indicated, sugars were highly accumulated in seed, e.g., glucose, fucose, erythritol and rhamnose; this was most apparent in the lupin species. On the other hand, our results indicate that *C. arietinum*, *P. vulgaris* and *L. culinaris* accumulate more fructose. Moreover, sucrose, maltose and xylose were mostly abundant in the IRC clade. In essence, the observed results provide insight into accumulation sites of precursor pools, such as the detected amino acids, upon grain filling and other developmental processes, which remain to date largely unknown. By discovering such sites, favorable paths can be integrated by the development of metabolomics-assisted breeding programs employing natural variation.

Lipidomic profiling revealed differences between genotypes and tissues

Lipids are a group of diverse compounds with crucial roles in all living organisms. They are found in membrane components, serve as storage sources for energy and are known signaling molecules in environmental stress responses (Wymann & Schneiter, 2008). Many studies have investigated lipids and their functions in plants (D'Ambrosio et al., 2017; Okazaki & Saito, 2014). In this study, we provide a comprehensive lipidome profile of the five investigated legume species across different tissues, which, to the best of our knowledge, have not been studied so far.

Using targeted LC-MS-based lipidomic analyses allowed the annotation of 923 lipid compounds, classified into the following chemical classes: TG (119), phosphatidylcholine (PC; 69), saccharolipid (SL; 59), diacylglycerol (DG; 39), phosphatidylinositol (PI; 22), monogalactosyldiacylglycerol (MGDG; 5), DGDG (22) and fatty acids (10), among others (Figure 1, Table S3). Additionally, we were able to detect and quantify nine pigments, namely, antheraxanthin, α-carotene, β-carotene, chlorophyll *a*, chlorophyll *b*, lutein, neoxanthin, pheophytin and violaxanthin, as well as multiple ceramide derivatives (Figure S1, Table S3). The results highlight tremendous variation in lipidome profiles between species and at the tissue level. The importance of different lipid classes has been reported in different plant tissues; for example, TGs highly accumulate in the endosperm of seeds and are used as carbon storage (reviewed in Xu and Shanklin (2016)), while MGDG and DGDG, both being major components of the photosynthetic membranes (reviewed in Dörmann and Benning (2002)), were found to be highly accumulated in leaves compared to root and seed tissues (Figure 3b). The results showed that PCs accumulate to a higher extent in leaves (Figure 3b). In seeds, the lipidome profiles display a wide range of variation between different species; for example, *L. albus* displayed high levels of several lipid classes, such as TGs and DGs (Figure S1). From the nutritional point of view, legumes



are rich in unsaturated fatty acids, composed mainly of oleic acid (C18:1), as well as essential polyunsaturated fatty acids like linoleic acid (C18:2) and α-linolenic acid (C18:3) (Khrisanapant et al., 2019). It has been shown that unsaturated fatty acids account for up to 63–85% of the total lipid composition (Khrisanapant et al., 2019). Our results indicate

Figure 3. Primary metabolic signatures. (a) Changes in major primary metabolites across tissues and species, from glycolysis to the tricarboxylic acid (TCA) cycle and further amino acid biosynthesis. For detected metabolites, the relative abundances are depicted in the pathway. (b) Heatmap illustrating the relative abundances of sugars, amino acids, organic acids and selected lipids (reporting level A) across species and tissues. Samples were hierarchically clustered with n_c = three clusters. (c) Relative abundances of free fatty acids in the seeds of different species. Samples were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) and genotypes (n_g = three to five) for the given species and tissues. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low).

that the two lupin species accumulate high levels of free fatty acids (including unsaturated long-chain [>18C] and un-/saturated short-chain [≤18C] fatty acids) as well as certain TGs, which are predominant in the seeds, while phospholipids and other TGs are highly present in *P. vulgaris*. Moreover, unsaturated long-chain free fatty acids showed higher abundance in *L. albus* and reduced levels in *L. culinaris* (Figure 3c). By contrast, the levels of saturated fatty acids like palmitic acid (C16:0) and stearic acid (C18:0) were marginal in *C. arietinum*, yielding in a low ratio of saturated to unsaturated fatty acids. This result is in agreement with the results of Khrisanapant et al. (2019), who showed that the ratio of saturated to unsaturated fatty acids in chickpea was 1:6.

Diversity and biosynthesis of specialized secondary metabolites in legumes

Legumes, similar to other plant species, synthesize a large diversity of specialized (or secondary) compounds that constitute the majority of their metabolomes. These compounds can be roughly divided into three main chemical classes, namely, alkaloids, polyphenolics and terpenoids. Each of these classes is comprised of many thousands of different individual compounds due to multiple and versatile decorations of a common skeleton by biochemical reactions, such as hydroxylation, glycosylation and methylation (reviewed in Alseekh et al. (2020)). Whilst considerable efforts have been made to identify and quantify these diverse compounds in model plants such as Arabidopsis, rice, maize and tomato (Butelli et al., 2008; Chen et al., 2013; Richter et al., 2016; Saito et al., 2013; Tohge et al., 2016, 2020) as well as other major crop species, a comprehensive characterization of specialized metabolites in legumes remains elusive. Here we were able to detect and quantify 2283 metabolites using polar LC-MS analysis. Given the majority of annotated specialized metabolites are classified as isoflavonoids, triterpene saponins and QAs, we will also describe our data in this manner in the following sections.

Iso-/flavonoid diversification

Iso-/flavonoids are among the most well-studied compound classes across the plant kingdom and have been reported to be involved in various aspects of plant ecophysiology, including pigmentation and plant (a)biotic interactions, such as herbivory, pollination, nodulation, UV radiation and high levels of salt and chilling conditions (Schulz et al., 2016; Tanaka et al., 2008; Tohge et al., 2016, 2018). Recently, polyphenolic compounds have also begun to be regarded as phytonutrients as they were reported to prevent cardiovascular diseases, diabetes and obesity (Martin et al., 2011; Tohge & Fernie, 2017). Among this class, compounds of the isoflavonoid subclass are bioactive compounds exclusively found in Leguminosae, exhibiting a wide range of decorative modifications. Historically, these compounds were initially seen as anti-nutritional, but they underwent a renaissance in the 1960s and 1987 when phytoestrogen and the tyrosine kinase inhibitory activity of the major isoflavonoid genistein were serendipitously demonstrated (Akiyama et al., 1987; Folman & Pope, 1966). Several modifications, such as hydroxylation, methylation, prenylation and phenylacylation, can occur (Alseekh et al., 2020; Tohge et al., 2017). However, studies assessing metabolic divergence at the tribe and tissue levels are rare, let alone reconstruction of their biochemical pathways.

Here, we were able to detect and quantify 342 iso-/flavonoid glycosides, including well-known previously reported aglycones that according to their MS^2 spectra vary in their degree of hydroxylation, methylation, prenylation and phenylacylation (Figure 4a, Table S1). In addition, we reconstructed the main biosynthetic pathways of these specific compounds and indicated the key major enzymes, such as flavonoid hydroxylase (FH), *O*-methyltransferases (OMTs), (aromatic) prenyltransferases ([A]PTs) and flavonoid phenylacyltransferases (FPTs) (Figure 4). Our results showed that various iso-/flavonoid glycosides were observed, such as the kaempferol aglycone decorated with pentaglycosides (Figure 4). Among the glycosylated flavonoids, apigenin diglycosides were one of the most common secondary metabolites found in Leguminosae and found to be highly abundant in *L. albus* leaves and seeds and in *L. mutabilis* roots. Further, accumulation of kaempferol and quercetin tetraglycoside was predominant in *L. culinaris*, particularly in seed tissue.

In addition, we reported several prenylated isoflavonoids in lupins. The occurrence of isoflavonoid prenylation by APTs has been described in Leguminosae (Akashi et al., 2009; Sasaki et al., 2008; Shen et al., 2012). Prenylation has been reported to improve the bioactivity of isoflavonoids by elevating the lipophilicity of the molecules and thereby increasing the affinity to target membrane proteins

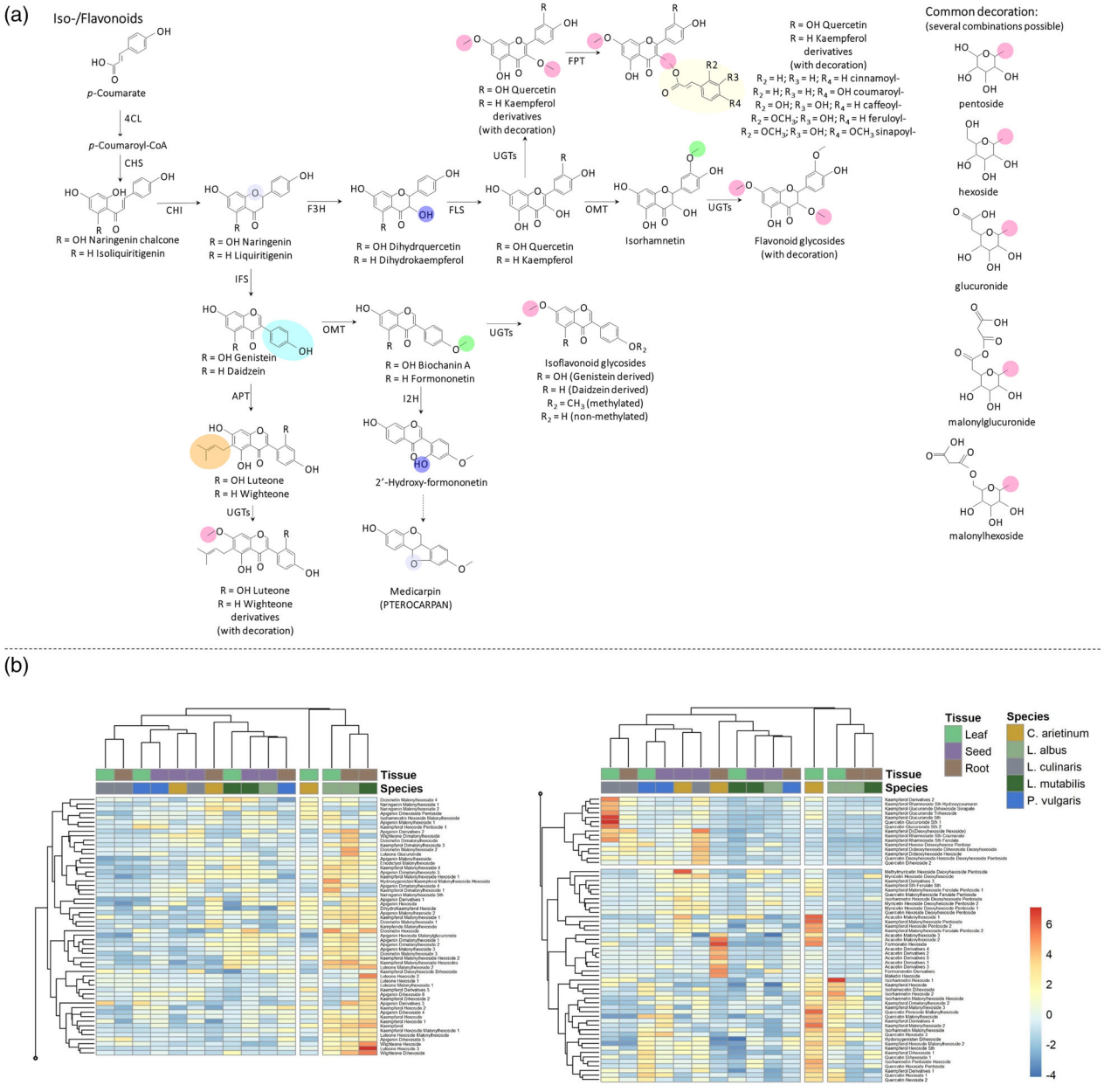


Figure 4. The putative biosynthetic pathways of iso-/flavonoids catalyzed by un-/characterized enzymatic steps in Leguminosae. Detected isoflavonoids are hierarchically clustered across samples and metabolites revealing molecular fingerprints. (a) Each arrow indicates an enzymatic step. While enzymatic steps with filled arrows are elucidated, dashed arrows are unknown and can reflect multiple enzymatic steps. Highlighted regions of the molecules represent the chemical modification upon each particular enzymatic step. Different colors reflect different modifications. Dark blue, green, orange, yellow, turquoise and magenta represent oxygenation, methylation, prenylation, phenylacylation, rearrangement of the phenolic ring and glycosylation, respectively. Common decoration sites represent single or multiple glycosylations with several potential glycosides at the highlighted position in magenta. CHS = chalcone synthase; CHI = chalcone isomerase; F3H = flavonoid-3′-hydroxylase; IFS = isoflavone synthase; APT = aromatic prenyltransferase; FPT = flavonol phenylacyltransferase; UGT = UDP-glycosyltransferase; OMT = *O*-methyltransferase. (b) Heatmap showing the relative abundances of detected flavonoids in species and tissues. Samples and metabolites were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) and genotypes (n_g = three to five) for the given species and tissues. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low).

(Shen et al., 2012). Further, anti-microbial, antioxidant, immunomodulatory and anti-cancer activities were previously reported (Venturelli et al., 2016). In this study, the detected prenylated isoflavonoid glycosides were mainly derived from luteone and wighteone, both showcasing predominant accumulation in the lupin species (Table S4). However, there was no clear separation between *L. mutabilis* and *L. albus* since both aglycone-containing

glycosides were present with particularly higher abundance in root tissue (Table S4; Figure 4). Specifically, the levels of luteone hexoside and wighteone dihexosides were higher in *L. mutabilis* roots, while luteone glucuronide and wighteone dimalonylhexoside, a hexose ester, were mostly abundant in *L. albus*. Previously, the *LaPT1* gene, encoding an enzyme catalyzing this particular modification, was identified in *L. albus*, and its function was validated in *Medicago truncatula*, leading to the biosynthesis of novel prenylated compounds in *M. truncatula*, highlighting potential strategies in improving crops (Shen et al., 2012).

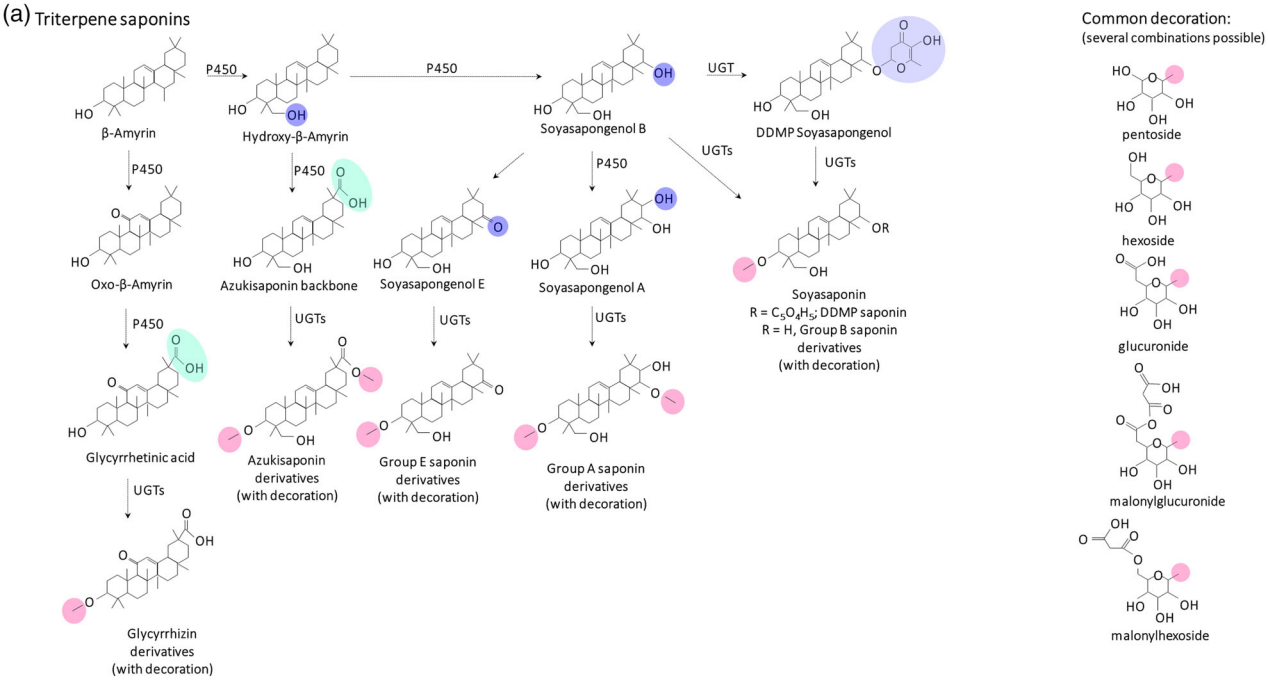
Further, our data revealed that lupins and chickpea produce unique and specific methoxylated iso-/flavonoid glycosides in comparison to the other species, indicating either a gain of function in the genisteae and cicereae clades or a loss of function in the fabeae and phaseoleae clades. That said, based on our assumption, the gain of function in genisteae for prenylated isoflavonoids seems highly likely to be the case here. However, deeper analysis integrating the sequence data from pan-genomes and metabolite profiles will probably enable greater insight into their evolutionary origin (Bellucci et al., 2021).

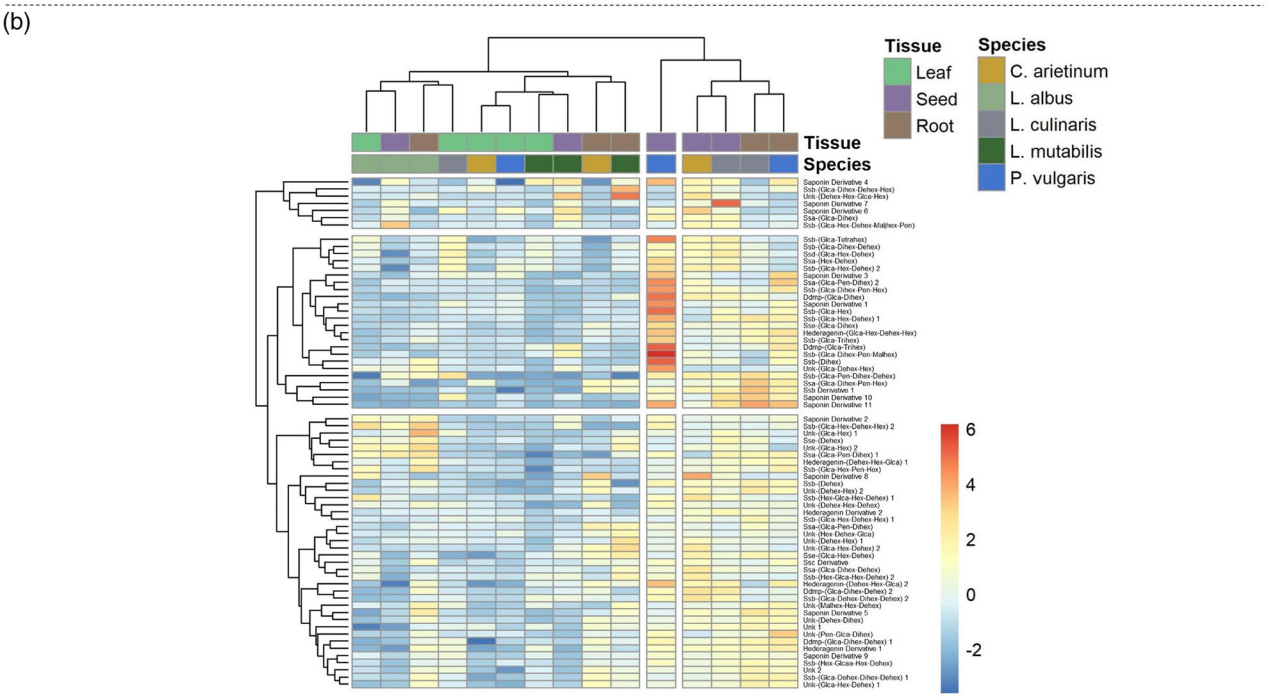
Furthermore, in this study, we showed that methoxylated iso-/flavonoids accumulate largely in leaves and roots, while the state of glycosylation varies depending on the tissue (Figure 4b). For instance, both acacetin and diosmetin aglycone are predominantly observed in the roots, while the additional hydroxylated form, isorhamnetin, accumulates markedly in leaves (Figure 4). In addition, flavonoid glycosides can be further acylated with phenylpropanoids, such as coumarate, ferulate and sinapate, by FPTs. Our results indicated that most of the phenylacylated compounds were found in the cicereae and fabeae, indicating a potential gain of function in the IRL clade, where cicereae and fabeae are divergent from faboideae (Figure 1). At the tissue level, these compounds mainly accumulate in leaves and seeds. Phenylacylation mostly occurs on kaempferol glycosides, with the exception of a single quercetin glycoside, suggesting that the enzyme catalyzing this reaction is considerably more specific than the recently characterized flavonol phenyltransferase 2 in *Arabidopsis thaliana* (Tohge et al., 2016). While methoxylated iso-/flavonoids have been reported to exert health-beneficial, anti-cancer and anti-inflammatory effects (Walle, 2007), phenylacylated flavonoids are less studied and it has been suggested that these are involved in UV protection, as phenylacylation enhances ultraviolet absorbance and antioxidant activity (Tohge et al., 2018). Besides these conventional approaches, the functional significance of the formation of flavonoid metabolons was highlighted recently, regulating their underlying fluxes and impacting flavonoid chemodiversity (Nakayama et al., 2019). Hence, altering the flavonoid composition provides an opportunity to improve their abundance for breeding programs by integrating genetic resources and identifying the underlying regulators. Especially their high accumulation in leaves makes these compounds of high value for potential fodder application as well as programs to improve overall plant resilience.

Oleanane saponin divergence

Triterpene saponins are major compounds in Leguminosae leading, from an anthropocentric view, to the bitter taste, which renders the majority of these compounds undesirable. They belong to the most diverse chemical class on earth. However, through co-evolution, these compounds play roles in several ecological interactions, such as defense against disease and herbivory (Mugford & Osbourn, 2013). Overall, several functions of those compounds have been described (see Zhou and Pichersky (2020)). Among the C-30 terpenoids, triterpene saponins occur in most tribes of Leguminosae and have been extensively studied (Rupasinghe et al., 2003; Sawai & Saito, 2011; Sayama et al., 2012; Sundaramoorthy et al., 2018; Yano et al., 2017). The majority found in Leguminosae are oleanane-triterpene saponin-derived, while lupane-triterpene saponins were also reported in the *genisteae* (Vincken et al., 2007). Biosynthetically, squalene is synthesized by squalene synthase/cyclase, followed by epoxygenation to 2,3-oxidosqualene by squalene epoxidase, from which oleananes and lupanes are derived. Downstream, P450-dependent monooxygenases and UDP-glycosyltransferases (UGTs) lead to oxidation and further glycosylation of the saponin aglycone, respectively.

In the current study, we were able to annotate 155 saponins (Table S1); while for several derivatives, we detected the main aglycone fragments but not the extent of subsequent decoration patterns. The majority of annotated saponins are part of the soyasapongenol subclass, with a large number of annotated compounds classified as group B saponins (Table S1). The soyasapogenol derivatives belong to the oleananes and are derived from β-amyrin, which is synthesized by the enzymatic reaction catalyzed by β-amyrin synthase (BAS1) (Jo et al., 2017; Takagi et al., 2011; Vernoud et al., 2021). In the case of soyasapogenol B, these metabolites are derived from the hydroxylation of β-amyrin at the C-22 position, while an additional hydroxylation at the C-21 position leads to the generation of group A saponins (Sawai & Saito, 2011; Sundaramoorthy et al., 2018; Yano et al., 2017). Metabolite profiling showed that soyasapongenols are detected in all species, with lower abundance found in the leaves compared to other tissues, indicating that the major catalytic steps for the saponin biosynthetic pathways are conserved across the species studied here. On the other hand, the abundance of saponins was predominant in seeds, with *P. vulgaris* and lupin seeds having the highest and lowest abundance, respectively (Figure 5). Recent reports indicate





the role of a novel UGT involved in the production of 2,3-dihydro-2,5-dihydroxy-6-methyl-4H-pyran-4-one (DDMP) saponin by binding DDMP at the soyasapogenol B C-22 hydroxylation side upon the release of water (Sundaramoorthy et al., 2019). To this end, we were able to detect several DDMP saponins among *P. vulgaris*, *L. culinaris* and *C. arietinum*. As stated above, triterpene saponins are associated with several ecological roles besides the bitter taste of pulses; however, it is believed that the aftertaste and bitterness of legume seeds originate from group A saponins (Rupasinghe et al., 2003), while DDMP saponins are less bitter and have stronger health-beneficial effects (Sayama et al., 2012), such as radical scavenging properties, anti-mutagenic activity (Berhow et al., 2002), inhibitory effects on HIV infection *in vitro* (Nakashima et al., 1989) and prevention of colon cancer proliferation (Tsai

Figure 5. The putative biosynthetic pathways of oleanane triterpene saponins catalyzed by un-/characterized enzymatic steps in Leguminosae. Detected saponins are hierarchically clustered across samples and metabolites revealing molecular fingerprints. (a) Each arrow indicates an enzymatic step. While enzymatic steps with filled arrows are elucidated, dashed arrows are unknown and can reflect multiple enzymatic steps. Highlighted regions of the molecules represent the chemical modification upon each particular enzymatic step. Different colors reflect different modifications. Dark blue, turquoise, magenta and blue represent oxygenation, carboxylation, glycosylation and DDMP modification, respectively. Common decoration sites represent single or multiple glycosylations with several potential glycosides at the highlighted position in magenta. P450 = cytochrome P450; UGT = UDP-glycosyltransferases. (b) Heatmap showing the relative abundances of detected saponins in species and tissues. Samples and metabolites were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) and genotypes (n_g = three to five) for the given species and tissues. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low). Hex = hexoside, Dehex = deoxyhexoside, Malhex = malonylhexoside, Glca = glucuronide, Pen = pentoside, Ssa = soyasaponin A, Ssb = soyasaponin B, Ssc = soyasaponin C, Sse = soyasaponin E, Unk = unknown.

et al., 2010). Based on these pharmacological effects and their roles in plant defense, saponins are highly interesting for grain legume improvement.

Diversity in quinolizidine alkaloids

Besides the more conserved metabolic classes described here, QAs are rare in the Leguminosae family. Some of those, namely lupin alkaloids, were first reported in *Lupinus* spp., as their nomenclature indicates, and are exclusively present in lupins and absent in the other investigated crop species (Boschin & Resta, 2013). These compounds are agronomically undesirable, due to their anti-nutritional properties, including a bitter taste and potential acute anti-cholinergic toxicity (Chain et al., 2019); as such, a threshold of 0.02% of dry seed weight for industrial production was introduced (Frick et al., 2017). The predominantly reported alkaloid derivatives in the Leguminosae family belong to the quinolizidine (QA), pyrrolizidine (PzA), indolizidine (IzA), piperidine (PiA), pyridine (PyA) and indole (InA) alkaloids (Wink, 2013). Further, according to Wink (2013), species in the genistoid clade (e.g., *L. albus* and *L. mutabilis*) synthesize QA and PiA alkaloids, whilst pyridine-derived alkaloids were reported for members of the IRLC clade (*C. arietinum* and *L. culinaris*). QAs, which are predominantly present in the *Lupinus* genus, are synthesized in the reaction catalyzed by LDC which forms cadaverine amine (Bunsupa, Katayama, et al., 2012), followed by the oxidation of the amine by a copper anime oxidase (Yang et al., 2017) yielding 5-aminopentanal (Bunsupa, Yamazaki, & Saito, 2012; Frick et al., 2017; Schäfer & Wink, 2009; Wink & Witte, 1985). Spontaneous cyclization forms piperideine, from which a diiminium cation and lupinine emerge (Figure 6a). The diiminium cation serves as a precursor for sparteine, multiflorine and lupanine, with the possibility of lupanine and multiflorine hydroxylation to 13a-hydroxylupanine and 13a-hydroxymultiflorine, respectively (Golebiewski & Spenser, 1988; Hemscheidt & Spenser, 1987). The hydroxylated sites of QA structures can undergo esterification with several acyl groups (Figure 6a), mainly tiglic acid, by hydroxylupanine/hydroxymultiflorine *O*-tigloyltransferases (HMTs/HLTs) (Suzuki et al., 1994). However, several other esters can be formed, for example, the one reported by acylation of the hydroxylupanine and hydroxymultiflorine backbone with angelic acid, benzoic acid, coumaric acid and cinnamic acid, as indicated in Figure 6 (Wink & Hartmann, 1982). Additionally, dihydroxylation of lupanine can result in diacylation with the previously mentioned substituents.

Here we were able to annotate 51 QAs. Several crucial observations were made while investigating lupin. The first striking observation was an accumulation of dihydroxylation of lupanine in *L. mutabilis*, while *L. albus* lacks the dihydroxylated derivatives (Figure 6b). Similarly, cinnamoyl esters were specific to *L. mutabilis* across all tissues, mainly accumulating in seeds and leaves. In general, multiple QA esters were conserved in both *L. mutabilis* and *L. albus*, while for some *L. mutabilis* genotypes benzoyl-, methylbutanoyl- and valeroyl-oxylupanine were not detected in the seeds. However, the alkaloid albine was exclusively detected in *L. albus*. Hierarchical clustering was performed for each chemical class to identify groups of compounds and samples with similar underlying fingerprints. The number of clusters was set to three for both metabolites and samples. Hierarchical cluster 3 alkaloids were predominantly present in *L. mutabilis*, including the dihydroxylated lupanine backbones (Figure 6b). On the other hand, the cluster 2 alkaloids were present in both lupin species, with higher overall levels observed in *L. albus*. As mentioned above, some of the cluster 2 alkaloids showed lower levels in seeds compared to the other tissues in *L. mutabilis*, e.g., benzoyl-, methylbutanoyl- and valeroyl-oxylupanine, while in *L. albus* these are distributed similarly across tissues (Figure 6b). Finally, cluster 1 alkaloids accumulate mainly in *L. albus* and *L. mutabilis* across tissues with the highest accumulation observed in seeds, while cytisine and several other alkaloid derivatives are highly abundant in *L. culinaris* (Figure 6). Since grain seeds are valuable sources of proteins, lipids and fibers, breeding programs aim to reduce the alkaloid abundance in seeds, while maintaining them at high levels in leaves and roots to reduce herbivory due to their insecticidal nature (Frick et al., 2017). However, depending on the potential application of leaves, these might also be considered undesirable. It has been shown that QAs are not only important from an agricultural point of view, but also

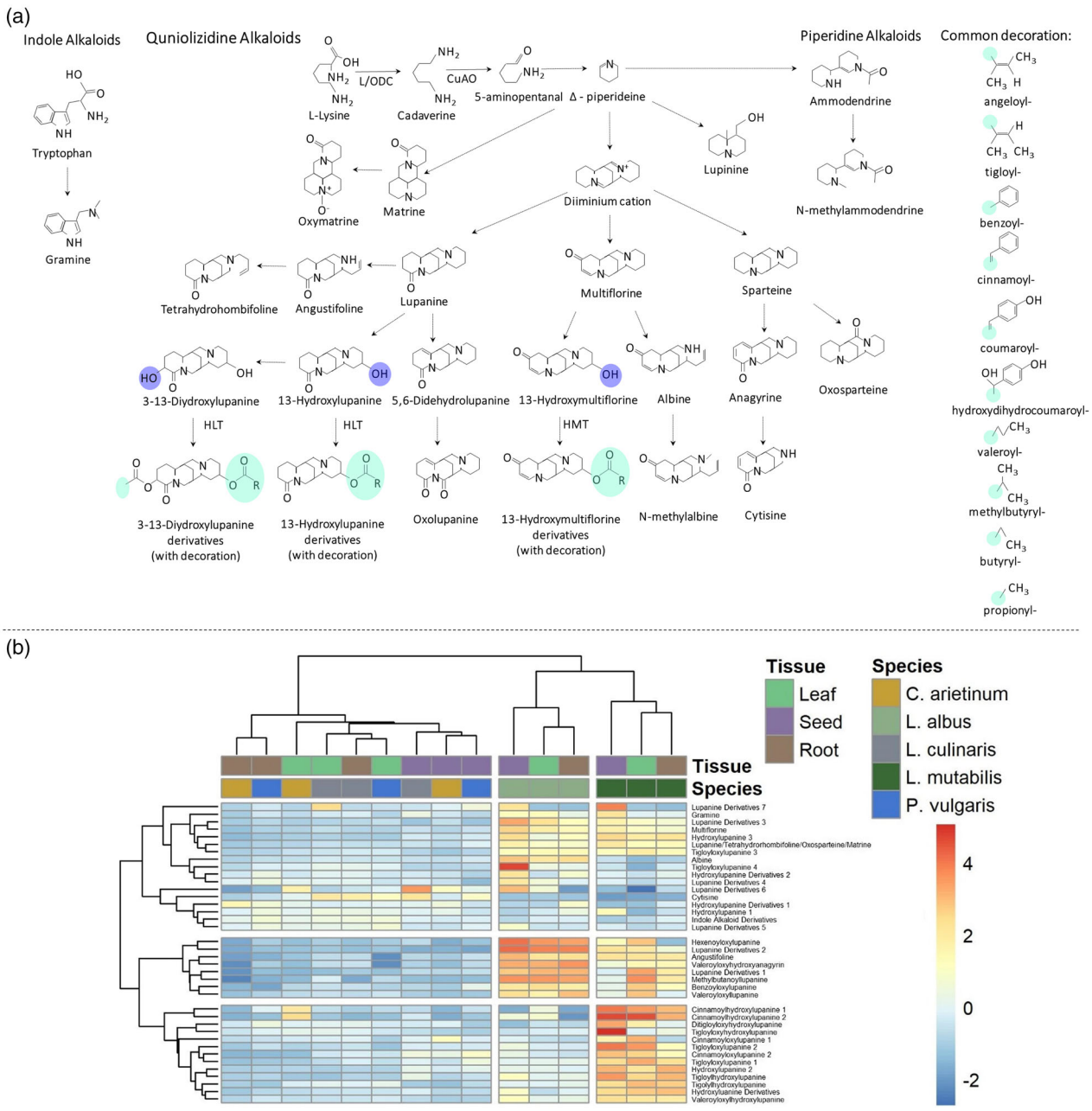


Figure 6. The putative biosynthetic pathways of alkaloids catalyzed by un-/characterized enzymatic steps in Leguminosae. Detected alkaloids are hierarchically clustered across samples and metabolites revealing molecular fingerprints. (a) Each arrow indicates an enzymatic step. While enzymatic steps with filled arrows are elucidated, dashed arrows are unknown and can reflect multiple enzymatic steps. Highlighted regions of the molecules represent the chemical modification upon each particular enzymatic step. Different colors reflect different modifications. Dark blue and turquoise represent oxygenation and acylation, respectively. The common decoration reflects possible acylation partners, which are esterified at the highlighted position in turquoise. L/ODC = lysine decarboxylase, CuAO = copper amine oxidase, HMT = hydroxymultiflorine *O*-tigloyltransferase, HLT = hydroxylupanine *O*-tigloyltransferase. (b) Heatmap showing the relative abundances of detected alkaloids in species and tissues. Samples and metabolites were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) and genotypes (n_g = three to five) for the given species and tissues. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low).

important in pharmacological aspects (reviewed in Wang et al. (2019)). A recent review on the known initial enzymatic steps of QA synthesis, e.g., CAO-dependent oxidation of cadaverine and the dimerization of piperideine, highlights the importance of identifying the functional genes (Mancinotti et al., 2022). Further, based on chemical modification and rearrangement, the authors proposed potential enzymatic steps in QA biosynthesis (Mancinotti

et al., 2022). In addition, the first enzymatic reaction of LDC was proven to lead to chemodiversity of nitrogen-containing metabolites, as overexpression of *LaLDC* in *A. thaliana* resulted in the detection of multiple novel cadaverine-derived metabolites (Shimizu et al., 2019). Kroc et al. (2019) reported an APETALA2/ethylene response factor, a transcription factor regulating QA composition; however, genes encoding catalytic enzymes in the biosynthetic pathway of QA mostly remain unelucidated in the *Lupinus* genus.

Intraspecies variation in metabolic composition

In addition to interspecies and tissue comparison described above, we also investigated the metabolic variation between the different genotypes from the same species. Our data showed that a quite high metabolic variation was observed within some of the genotypes of the same species (see Table S4); for example, the QA distribution highlights clear genotypic variation for two dihydroxylated lupanin derivatives in *L. mutabilis* seeds (Figure S2). For the triterpene saponins, intraspecific variation was observed in *P. vulgaris*; for example, large variation in soyasapongenol levels was reported between *P. vulgaris* genotypes (Figure S3 or Table S4). These variations likely originate from the two distinct domestication events, which took place in the Andes and Mesoamerica (Bellucci et al., 2014, 2023). Another example of metabolic diversity was observed between the domestication origin of desi and kabuli genotypes in chickpea; for example, the levels of two myrecetin glycosides in seeds and kaempferol derivatives in leaves were associated with the desi origin (Figure S4) (Rocchetti et al., 2020; Varshney et al., 2019). Overall, the observed intraspecific diversity enables the investigation and further uncovering of genes involved in the regulation of each individual chemical class by screening large populations diverging in their levels of metabolites of interest. Indeed, metabolite-based genome-wide association studies (mGWAS) on the above investigated species have identified enormous loci and genetic factors regulating a wide range of metabolites.

In silico metabolic prediction

In order to extend our list of annotated metabolites, we applied here a mathematical approach to assign putative identities to unknown detected molecules. Recently, computational approaches to annotate metabolites have made considerable advances, especially the appearance of feature-based molecular networking (FBMN; Nothias et al. (2020)) and artificial intelligence-based prediction of fragmentation trees in SIRIUS (Dührkop et al., 2019). Therefore, we implemented the approach as described in Tripathi et al. (2021) to compare the performance of the prediction model of SIRIUS with that of our conventional metabolite annotation and to ultimately increase the number of annotated compounds. To avoid mis-alignment of fragments to parent ions, DDA-MS^2 spectra were used on separate tissue-wise pooled samples and aligned with the feature intensities for each sample in MS^1. The prediction resulted in the identification of 663 $DDMS^2$ spectra, while an additional 430 mass spectra were unassigned (Table S6). For instance, the chemical superclass assignments 'Alkaloid and derivatives', 'Phenylpropanoids and polyketides' and 'Lipids and lipid-like molecules' harboring the chemical classes described above were well identified and highlight major diversity with species and tissue specificity (Figure 7). Briefly, the alkaloid branch shows lupin specificity with respective intraclade chemodiversity, whereas the triterpene saponins predominantly accumulate in seed and root tissues in *P. vulgaris*. Unfortunately, due to the restriction of the parent ion to 1200 *m*/*z*, multiple triterpene saponins, and other larger molecules, were out of the prediction range and thereby biased the analysis. However, novel observations to previously described metabolic distribution were indicated, e.g., high levels of 'amino acids, peptides and analogues' in the species *L. culinaris* and *P. vulgaris*, especially in seeds. The table of predicted identities is provided in Table S6, in parallel with the output files, which can be used to generate and investigate the hierarchical clusters by uploading to https://view.qiime2.org/.

CONCLUSION

In this article, we provide an in-depth chemical characterization of divergent Leguminosae tribes at the tissue level for the scientific community. For several metabolites, high intraspecies variation was observed across five legume species (Table S4). This variation mainly traces back to the domestication events in the assessed crop species; e.g., two distinct domestication events have been reported for *P. vulgaris* (Andean and Mesoamerican), as well as the presence of two distinct populations, desi and cabuli, for *C. arietinum*. Computational metabolomics approaches, such as *in silico* metabolite prediction, perform well to assign chemical classes as shown for major compound classes (Figure 7). However, the prediction lacks precision in assigning the decorations (glycosylation number and sequence). Comparing the sheer number of detected metabolites across the species with publicly available databases, such as KNApSAcK (Afendi et al., 2012; Shinbo et al., 2006), it becomes evident that most of the species addressed are well understudied (*L. mutabilis* = 15 metabolites; *L. culinaris* = 16 metabolites and *C. arietinum* = 43 metabolites), whereas the other two are partially more investigated in their chemical profiles (*L. albus* = 177 metabolites and *P. vulgaris* = 521 metabolites), leading to the necessity of broadening information on their chemodiversity and providing it for society.

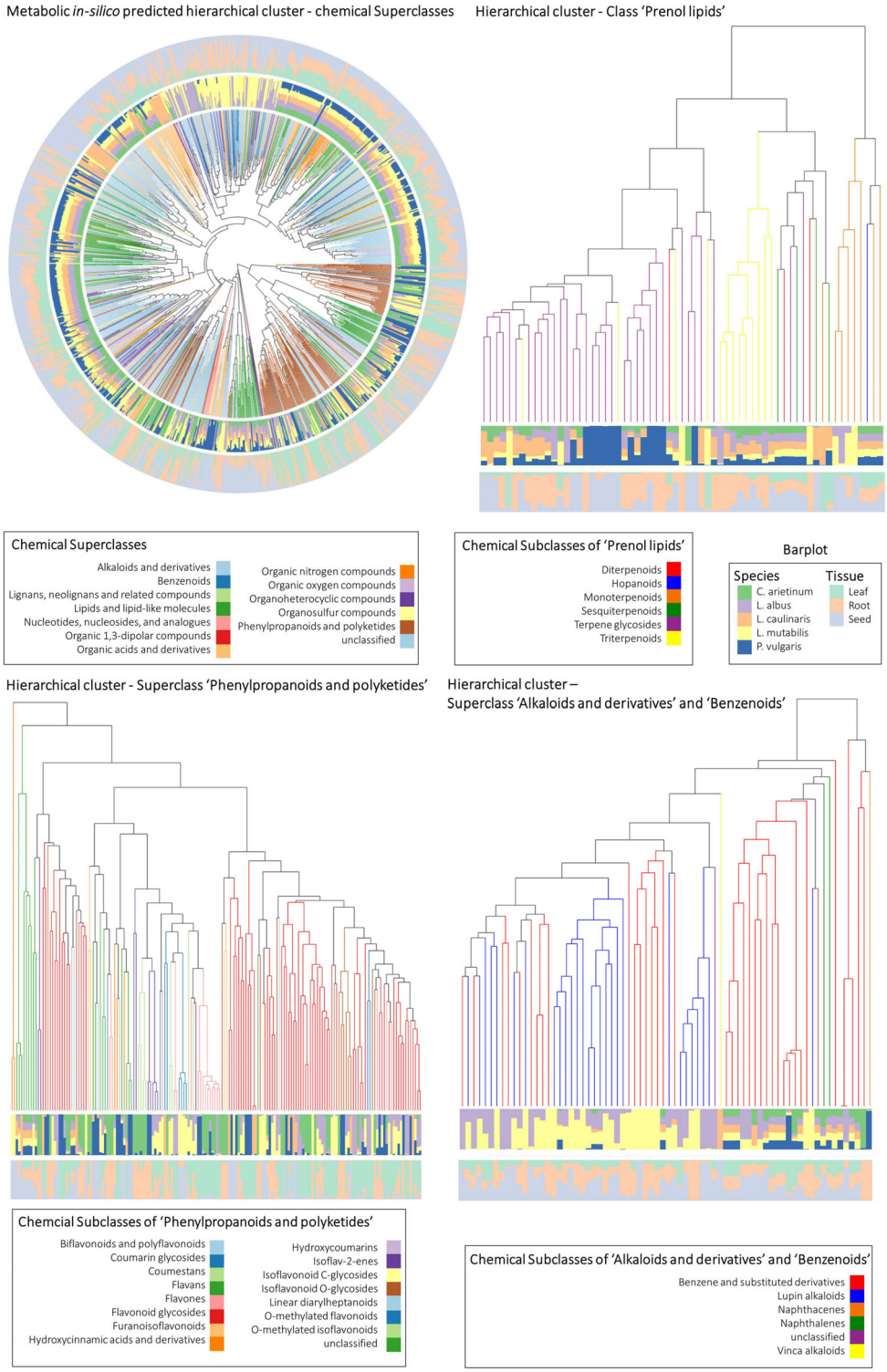


Figure 7. Chemical hierarchy of specialized metabolites in Leguminosae. The chemical phylogeny is generated by *in silico* prediction of detected metabolic features. The hierarchical trees represent predicted chemical structures with their assigned chemical superclasses and subclasses based on Classyfire. In-depth subclass prediction of the chemical superclasses 'Phenylpropanoids and polyketides', 'Alkaloids and derivatives' and 'Benzenoids' and the chemical class 'Prenyl lipids' with their distribution across species and tissues is given in the hierarchical trees. The inner barplot reflects the relative abundances species-wise, whereas the outer barplot reflects the relative abundances tissue-wise. The table of predicted metabolite identities is provided in Table S4.

Further, the majority of the biosynthetic pathways of the described specialized metabolites is to date unelucidated across the addressed species. As discussed above, each of those metabolite classes can have a detrimental impact on health, combatting hidden hunger and improving feed application; therefore, their role in crop improvement is of crucial importance. Combining metabolic databases, as demonstrated here, with the advances in genome sequencing and the creation of genetic resources will initiate a wave of gene discovery. A step forward would be the usage of the genetic resources as described in Bellucci et al. (2021), which would enable the elucidation of biochemical pathways, as high metabolic diversity is described in the Leguminosae family. Indeed, there are studies that highlight potential candidate genes based on transcription profiles and ortholog search from better elucidated species in Leguminosae, e.g., model organisms like *Glycine max* and *M. truncatula*; however, molecular validation of involved genes is rare. To this end, powerful functional genomics tools have been established in the past years to validate the biological function of identified candidate genes in the crops of interest using CRISPR/Cas-mediated genome editing and hairy root transformation (Alamillo et al., 2023). Several ecophysiological aspects have been assessed using natural variation, e.g., as shown for metabolism in other important crop species (Alseekh, Kostova, et al., 2021; Bulut et al., 2022; Luo, 2015; Oikawa et al., 2008; Perez-Fons et al., 2019; Sakurai, 2022; Sumner et al., 2015; Zhang et al., 2021). Finally, the integration of the metabolic fingerprints and mGWAS data with the creation of pan-genomes will provide clear evidence of divergence in the Leguminosae family as well as the evolution of gene clusters and families (Bellucci et al., 2021).

EXPERIMENTAL PROCEDURES

Plant material

A total of 22 accessions across five legume species (*P. vulgaris*, *L. culinaris*, *C. arietinum*, *L. albus* and *L. mutabilis*) and three tissues, i.e., root, leaf and seed (see Table 1), were collected and subjected to further analysis. In detail, fully developed mature leaves and rootstock of 4-week-old plants were harvested in the middle of the light period, snap-frozen in liquid N_2 and stored at −80°C until further analysis. Rootstocks were washed with deionized water prior to snap-freezing to minimize soil contamination. Seeds were processed in a dried state. Three to five genotypes each with four replicates were grown under controlled conditions (16 h/8 h day/night cycle; 26°C/20°C day/night and 70% relative humidity) in soil. Accessions of each species used in the study are compiled in Table 1.

Metabolite extraction and measurements

Metabolite extraction was performed as described in Giavalisco et al. (2011) and Salem et al. (2020) with some minor modifications. Briefly, 1 ml of pre-cooled (−20°C) methyl *tert*-butyl ether: methanol (3:1 v/v) was added to homogenized tissues. After, tubes were thoroughly vortexed for 1 min and then incubated on an orbital shaker (1000 rpm) for 15 min at 4°C followed by sonication for 15 min. For phase separation, a volume of 500 μl of water:

Table 1 List of chemically profiled accessions in this study

Species (Genotype)	Gene bank code	PGRFA DOI number	Donor Institute
Lens culinaris (G1)	ILL 213	10.18730/13K6E6	ICARDA
Lens culinaris (G2)	ILL 11557	10.18730/13 K691	ICARDA
Lens culinaris (G3)	PI 299351 LSP	10.18730/13K9V4	ICARDA
Lens culinaris (G4)	Val di Nevola	10.18730/13KFAZ	UNIBAS
Lens culinaris (G5)	Palazzo San Gervasio	10.18730/13KF7W	UNIBAS
Cicer arietinum (G1)	CIC 63	10.18730/Z8HN=	IPK
Cicer arietinum (G2)	Ares	10.18730/Z8M54	SAIS - seed company
Cicer arietinum (G3)	W6 26 221	10.18730/ZAB5S	USDA
Cicer arietinum (G4)	PI 533675	10.18730/Z8NPG	USDA
Cicer arietinum (G5)	W6 26 027	10.18730/ZAA7*	USDA
Lupinus albus (G1)	PL95530	10.18730/13N0SQ	IHAR-PIB
Lupinus albus (G2)	PL95125	10.18730/13MQFN	IHAR-PIB
Lupinus albus (G3)	PL95119	-	IHAR-PIB
Lupinus albus (G4)	PL95513	-	Poznan Plant Breeders Ltd.
Lupinus mutabilis (G1)	LUP 6279	-	IPK
Lupinus mutabilis (G2)	LUP 6204	-	IPK
Lupinus mutabilis (G3)	PL98904	10.18730/13M73F	IHAR-PIB
Phaseolus vulgaris (G1)	PI352755	10.18730/H7Q18	USDA
Phaseolus vulgaris (G2)	PHA2241	10.18730/H7Z15	IPK
Phaseolus vulgaris (G3)	PI312092	10.18730/H7RJM	USDA
Phaseolus vulgaris (G4)	PI313590	10.18730/H7RQS	USDA
Phaseolus vulgaris (G5)	PHA3677	10.18730/H80SR	IPK

PGRFA, Plant Genetic Resources for Food and Agriculture; ICARDA, International Center for Agricultural Research in the Dry Areas; UNIBAS, University of Basilicata; IPK, Leibniz Institute of Plant Genetics and Crop Plant Research; USDA, United States Department of Agriculture; IHAR-PIB, Plant Breeding and Acclimatization Institute.

methanol (3:1 v/v) was added to each tube and the samples were again thoroughly vortexed for 1 min. After that, the samples were centrifuged at 13 300 *g* for 5 min. The semi-polar phase was used for specialized and primary metabolite analysis, while the organic phase was used for lipid analysis. For specialized metabolites, dried polar aliquots were resuspended in water:methanol (1:1 v/v). Semi-polar metabolite analysis was performed on a Thermo Q Exactive Focus (Thermo Fisher Scientific, Waltham, MA, USA) with a reverse phase C_{18} column held at 40°C with a flow rate of 400 μl min^{-1} with gradual changes of eluent A (water) and eluent B (acetonitrile), both stabilized with 0.1% formic acid. Mass spectra were acquired in full scan in a range of 100–1500 *m/z* with data-independent acquisition (DIA) of MS^e having the high-energy collisional dissociation (HCD) energy set at 30 eV in positive mode. Additionally, samples were guided with pooled tissue-wise samples of DIA in positive and negative mode as well as in data-dependent acquisition (DDA) MS^2 for specifying parental ion association and fragmentation of specialized metabolites.

Dried extracts were resuspended in acetonitrile:isopropanol (7:3) for lipid analysis, and derivatization was performed according to Lisec et al. (2006) for primary metabolite analysis. In brief, derivatization was performed for 120 min at 37°C in 40 μl of 20 mg ml^{-1} methoxyamine hydrochloride (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany, cat. no. 593-56-6) in pyridine (Merck KGAA, Darmstadt, Germany, cat. no. 110-86-1), followed by a 30-min treatment at 37 °C with 70 μl of trimethylsilyl-*N*-methyl trifluoroacetamide (MSTFA; Macherey-Nagel GmbH & Co.KG, Düren, Germany, Ref. no. 701270.510). An autosampler Gerstel Multi-Purpose system (Gerstel GmbH & Co.KG, Mülheim an der Ruhr, Germany) was used to inject 1 μl of the samples in splitless mode into a chromatograph coupled to a time-of-flight mass spectrometer system (Leco Pegasus HT TOF-MS; LECO Corporation, St. Joseph, MI, USA). Helium was used as carrier gas at a constant flow rate of 2 ml sec^{-1} and gas chromatography was performed on a 30-m DB-35 column (capillary column, 30 m length, 0.32 mm inner diameter, 0.25 μm film thickness, PN: G42, Agilent Technologies Deutschland GmbH, Waldbronn, Germany). The injection temperature was 230°C and the transfer line and ion source temperatures were set to 250°C. The initial temperature of the oven (85°C) increased at a rate of 15°C min^{-1} up to a final temperature of 360°C. After a solvent delay of 180 sec, mass spectra were recorded at 20 scans sec^{-1} with 70–600 *m/z* scanning range.

Lipid analysis was performed on an Orbitrap LC-MS system (Exactive, ThermoScientific). Metabolite separation was performed on a reversed-phase C_{18} column with gradual changes in eluent composition, namely, water and acetonitrile:2-propanol (7:3 v/v), both stabilized in 0.1% acetic acid, held at 60°C running with a flow rate of 400 μl min^{-1}. Mass spectra were acquired in full MS scan in positive ionization mode with a mass range of 150–1500 *m/z*. For determining free fatty acids in seeds, lipid extracts were additionally run in negative ionization mode. A more in-depth description of each analytical platform can be found in Bulut et al. (2021).

Chromatogram & mass spectra processing

LC-MS full scan data and GC-MS data were processed using MS Refiner (Expressionist 14.0). In brief, this includes (i) chromatogram alignment based on an allowed retention time shift, (ii) peak detection based on the reoccurrence of certain masses over several scans and their overall shape, (iii) noise filtering based on a set noise level, (iv) isotope clustering and (v) singleton filtering.

DDA-MS^2 data for the pooled tissue-wise samples were processed using MzMine2 (Pluskal et al., 2010). After spectral alignment and peak detection, all features having associated MS^2 spectra were exported compatible with SIRIUS (Dührkop et al., 2019) and aligned to the full scan acquired data processed with MS Refiner. Additionally, DDA-MS^2 and non-polar LC chromatograms of each species per tissue were processed using MS-DIAL (Tsugawa et al., 2020). For GC-MS chromatograms, further processing was conducted using Xcalibur (ThermoScientific) and Pegasus (LECO).

Compound annotation

Specialized metabolite annotation was performed mainly by reconstructing the fragmentation pattern consolidating neutral losses, presence of adducts and ionization mode as well as homo- and heterolytic cleavage. In addition, fragmentation pattern searches were performed using ReSpect (http://spectra.psc.riken.jp/menta.cgi/respect/search/fragment). Further, non-targeted compound identification was performed using the authentic standards library in positive ion mode detection consisting of 16 481 unique compounds using CompMS (CompMS | MS-DIAL (riken.jp)). Additionally, the negative ion mode MS^1 dataset was likewise processed using MSDIAL to assign negative ionized fragments to the detected positive ions. For the *in silico* prediction of metabolite identities, the qemistree (Tripathi et al., 2021) pipeline was used. Briefly, the *in silico* prediction generates fragmentation trees for each molecular formula obtained in MS^1 using SIRIUS (Dührkop et al., 2019) by first ranking the molecular formula in MS^1 by Zodiac (Ludwig et al., 2020) and then predicting molecular substructures of each parent ion using CSI:FingerID (Dührkop et al., 2015). Qemistree restricts the generation of fragmentation trees of parent ions to 1200 *m/z*. Molecules are afterwards classified by assigning to the Classyfire (Djoumbou Feunang et al., 2016) chemical taxonomy. Finally, q2-empress was used to generate a molecular hierarchy, which then was visualized and exported using Qiime2 Viewer. The general qemistree output illustrates the presence and/or absence of chemical structures across samples. Instead of indicating the presence or absence of predicted metabolites in the phylogeny, the bar plots were modified to represent the relative abundances of each species/tissue over the sum of all samples.

For annotation of metabolites derived from GC-MS, an in house library generated by standards was used for sugar, amino acid and organic acid annotations. Further, a broader approach for assigning metabolite identity was performed in aligning electron ionization (EI)-induced fragments to the best hits for the deconvoluted mass spectra using in house, publicly available libraries, including Golm Metabolome Database (Kopka et al., 2005), as well as the National Institute of Standards and Technology (NIST) library. Lipid annotation was mainly performed by in house library search based on full scan MS^1, consisting of standalone standards, and stepwise changes in polarity with increasing unsaturation degrees as described in Hummel et al. (2011), with the additional identity assignment from the lipidome atlas of MS-DIAL (Tsugawa et al., 2020) with a mass error up to 5 ppm. All metabolites are reported following the recently updated reporting standards for metabolomics (Alseekh Aharoni, et al., 2021).

AUTHOR CONTRIBUTIONS

SA designed the experiments. MB, RW and SA performed the lab work. MB performed data analysis. MB, SA and ARF wrote the manuscript. All authors revised the manuscript and gave their input.

ACKNOWLEDGMENTS

MB acknowledges the funding by the IMPRS-MolPlant. EBi, EBe, MK, TG, KS, RP, ARF and SA acknowledge the financial support from the EU Horizon 2020 research and innovation Programme, project INCREASE (grant agreement No. 862862). AFR and SA acknowledge the financial support by the EU Horizon 2020 Research and Innovation Programme, project PlantaSYST (SGA-CSA No. 739582 under FPA No. 664620). Figures 1 and 3 were created using BioRender.com. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Data S1. QZV file of the predicted identities for viewing the chemical taxonomies and relative abundances across species in https://view.qiime2.org/.

Data S2. QZV file of the predicted identities for viewing the chemical taxonomies and relative abundances across tissues in https://view.qiime2.org/.

Figure S1. Lipidome diversity. Heatmap of all lipids detected across species and tissues, respectively. Samples and metabolites were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) and genotypes (n_g = three to five) for the given species and tissues. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low).

Figure S2. Alkaloid diversity in *Lupinus* spp. Heatmap of all detected alkaloids across all individual accessions in the lupin species. Metabolites were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) for the given species, tissues and genotypes. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low).

Figure S3. Saponin diversity in *Phaseolus vulgaris*. Heatmap of all detected triterpene saponins across all individual accessions in *P. vulgaris*. Metabolites were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) for the given species, tissues and genotypes. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low).

Figure S4. Iso-/flavonoid diversity in *Cicer arietinum*. Heatmap of all detected iso-/flavonoids across all individual accessions in *C. arietinum*. Metabolites were hierarchically clustered with n_c = three clusters. Depicted intensities represent calculated means across all replicates (n_r = four) for the given species, tissues and genotypes. The color code indicates the relative trait value normalized by $\log_{10}$ transformation and pareto scaling (orange = high and blue = low).

Table S1. Metabolite annotation of specialized metabolites using HPLC-MS/MS.

Table S2. Metabolite annotations of derivatized metabolites using GC-MS.

Table S3. Metabolite annotation of lipids using HPLC-MS.

Table S4. Detected raw quantities for each metabolic feature across each sample.

Table S5. Loadings of each metabolite in the PCA.

Table S6. Predicted identities based on Sirius using the qemistree pipeline.

REFERENCES

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K. *et al.* (2012) KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant and Cell Physiology*, **53**, e1.

Akashi, T., Sasaki, K., Aoki, T., Ayabe, S.-i. & Yazaki, K. (2009) Molecular cloning and characterization of a cDNA for Pterocarpan 4-Dimethylallyltransferase catalyzing the key Prenylation step in the biosynthesis of Glyceollin, a soybean Phytoalexin. *Plant Physiology*, **149**, 683–693.

Akiyama, T., Ishida, J., Nakagawa, S., Ogawara, H., Watanabe, S.-i., Itoh, N. *et al.* (1987) Genistein, a specific inhibitor of tyrosine-specific protein kinases. *Journal of Biological Chemistry*, **262**, 5592–5595.

Alamillo, J.M., López, C.M., Martínez Rivas, F.J., Torralbo, F., Bulut, M. & Alseekh, S. (2023) Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein and hairy roots: a perfect match for gene functional analysis and crop improvement. *Current Opinion in Biotechnology*, **79**, 102876.

Alseekh, S., Aharoni, A., Brotman, Y., Contrepois, K., D'Auria, J., Ewald, J. *et al.* (2021) Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods*, **18**, 747–756.

Alseekh, S. & Fernie, A.R. (2018) Metabolomics 20 years on: what have we learned and what hurdles remain? *The Plant Journal*, **94**, 933–942.

Alseekh, S., Kostova, D., Bulut, M. & Fernie, A.R. (2021) Genome-wide association studies: assessing trait characteristics in model and crop plants. *Cellular and Molecular Life Sciences*, **78**, 5743–5754.

Alseekh, S., Perez de Souza, L., Benina, M. & Fernie, A.R. (2020) The style and substance of plant flavonoid decoration; towards defining both structure and function. *Phytochemistry*, **174**, 112347.

Bar, N., Korem, T., Weissbrod, O., Zeevi, D., Rothschild, D., Leviatan, S. *et al.* (2020) A reference map of potential determinants for the human serum metabolome. *Nature*, **588**, 135–140.

Bellucci, E., Benazzo, A., Xu, C., Bitocchi, E., Rodriguez, M., Alseekh, S. *et al.* (2023) Selection and adaptive introgression guided the complex evolutionary history of the European common bean. *Nature Communications*, **14**, 1908.

Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S. *et al.* (2014) Decreased nucleotide and expression diversity and modified Coexpression patterns characterize domestication in the common bean. *The Plant Cell*, **26**, 1901–1912.

Bellucci, E., Mario Aguilar, O., Alseekh, S., Bett, K., Brezeanu, C., Cook, D. *et al.* (2021) The INCREASE project: intelligent collections of food-legume genetic resources for European agrofood systems. *The Plant Journal*, **108**, 646–660.

Berhow, M.A., Cantrell, C.L., Duval, S.M., Dobbins, T.A., Maynes, J. & Vaughn, S.F. (2002) Analysis and quantitative determination of group B saponins in processed soybean products. *Phytochemical Analysis*, **13**, 343–348.

Bornet, F.R.J. (1994) Undigestible sugars in food products. *The American Journal of Clinical Nutrition*, **59**, 763S–769S.

Boschin, G. & Resta, D. (2013) Alkaloids Derived from Lysine: Quinolizidine (a Focus on Lupin Alkaloids). In: Ramawat, K.G. & Mérillon, J.-M. (Eds.) *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 381–403.

Brown, K.S. Jr & Trigo, J.R. (1995) The ecological activity of alkaloids. *The Alkaloids: Chemistry and Pharmacology*, **47**, 227–354.

Bulut, M., Alseekh, S. & Fernie, A.R. (2022) Natural variation of respiration-related traits in plants. *Plant Physiology*, **191**, 2120–2132.

Bulut, M., Fernie, A.R. & Alseekh, S. (2021) Large-scale multi-omics genome-wide association studies (Mo-GWAS): guidelines for sample preparation and normalization. *JoVE*, **173**, e62732.

Bunsupa, S., Katayama, K., Ikeura, E., Oikawa, A., Toyooka, K., Saito, K. *et al.* (2012) Lysine decarboxylase catalyzes the first step of Quinolizidine

alkaloid biosynthesis and coevolved with alkaloid production in Leguminosae. *The Plant Cell*, **24**, 1202–1216.

Bunsupa, S., **Yamazaki, M.** & **Saito, K.** (2012) Quinolizidine alkaloid biosynthesis: recent advances and future prospects. *Frontiers in Plant Science*, **3**. Available from: https://doi.org/10.3389/fpls.2012.00239

Butelli, E., **Titta, L.**, **Giorgio, M.**, **Mock, H.-P.**, **Matros, A.**, **Peterek, S.** ***et al.*** (2008) Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nature Biotechnology*, **26**, 1301–1308.

Buxton, D.R., **Russell, J.R.** & **Wedin, W.F.** (1987) Structural neutral sugars in legume and grass stems in relation to Digestibility1. *Crop Science*, **27**, 1285 cropsci1987.0011183X002700060038x.

Capper, B.S. (1990) The role of food legume straw and stubble in feeding livestock. In: Osman, A.E., Ibrahim, M.H. & Jones, M.A. (Eds.) *The Role of Legumes in the Farming Systems of the Mediterranean Areas: Proceedings of a Workshop on the Role of Legumes in the Farming Systems of the Mediterranean Areas UNDP/ICARDA, Tunis, June 20–24, 1988*. Dordrecht: Springer Netherlands, pp. 151–162.

Chain, E., **Panel, o.C.i.t.F.**, **Schrenk, D.**, **Bodin, L.**, **Chipman, J.K.**, **del Mazo, J.** ***et al.*** (2019) Scientific opinion on the risks for animal and human health related to the presence of quinolizidine alkaloids in feed and food, in particular in lupins and lupin-derived products. *EFSA Journal*, **17**, e05860.

Chen, W., **Gong, L.**, **Guo, Z.**, **Wang, W.**, **Zhang, H.**, **Liu, X.** ***et al.*** (2013) A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of Rice metabolomics. *Molecular Plant*, **6**, 1769–1780.

Cortinovis, G., **Oppermann, M.**, **Neumann, K.**, **Graner, A.**, **Gioia, T.**, **Marsella, M.** ***et al.*** (2021) Towards the development, maintenance, and standardized phenotypic characterization of single-seed-descent genetic resources for common bean. *Current Protocols*, **1**, e133.

Cubero-Leon, E., **Peñalver, R.** & **Maquet, A.** (2014) Review on metabolomics for food authentication. *Food Research International*, **60**, 95–107.

Czepiel, K., **Krajewski, P.**, **Wilczura, P.**, **Bielecka, P.**, **Święcicki, W.** & **Kroc, M.** (2021) Expression Profiles of Alkaloid-Related Genes across the Organs of Narrow-Leafed Lupin (Lupinus angustifolius L.) and in Response to Anthracnose Infection. *International Journal of Molecular Sciences*, **22**, 2676.

D'Ambrosio, J.M., **Couto, D.**, **Fabro, G.**, **Scuffi, D.**, **Lamattina, L.**, **Munnik, T.** ***et al.*** (2017) Phospholipase C2 affects MAMP-triggered immunity by modulating ROS production. *Plant Physiology*, **175**, 970–981.

Desmet, S., **Saeys, Y.**, **Verstaen, K.**, **Dauwe, R.**, **Kim, H.**, **Niculaes, C.** ***et al.*** (2021) Maize specialized metabolome networks reveal organ-preferential mixed glycosides. *Computational and Structural Biotechnology Journal*, **19**, 1127–1144.

Djoumbou Feunang, Y., **Eisner, R.**, **Knox, C.**, **Chepelev, L.**, **Hastings, J.**, **Owen, G.** ***et al.*** (2016) ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics*, **8**, 61.

Dörmann, P. & **Benning, C.** (2002) Galactolipids rule in seed plants. *Trends in Plant Science*, **7**, 112–118.

Drapal, M., **Enfissi, E.M.A.** & **Fraser, P.D.** (2022) The chemotype core collection of genus Nicotiana. *The Plant Journal*, **110**, 1516–1528.

Dührkop, K., **Fleischauer, M.**, **Ludwig, M.**, **Aksenov, A.A.**, **Melnik, A.V.**, **Meusel, M.** ***et al.*** (2019) SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, **16**, 299–302.

Dührkop, K., **Shen, H.**, **Meusel, M.**, **Rousu, J.** & **Böcker, S.** (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences*, **112**, 12580–12585.

Fernie, A.R., **Tadmor, Y.** & **Zamir, D.** (2006) Natural genetic variation for improving crop quality. *Current Opinion in Plant Biology*, **9**, 196–202.

Fiehn, O. (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology*, **48**, 155–171.

Folman, Y. & **Pope, G.S.** (1966) The interaction in the immature mouse of potent oestrogens with coumestrol, genistein and other uterovaginotrophic compounds of low potency. *Journal of Endocrinology*, **34**, 215–225.

Frick, K.M., **Kamphuis, L.G.**, **Siddique, K.H.M.**, **Singh, K.B.** & **Foley, R.C.** (2017) Quinolizidine alkaloid biosynthesis in Lupins and prospects for grain quality improvement. *Frontiers in Plant Science*, **8**. Available from: https://doi.org/10.3389/fpls.2017.00087

Giavalisco, P., **Li, Y.**, **Matthes, A.**, **Eckhardt, A.**, **Hubberten, H.-M.**, **Hesse, H.** ***et al.*** (2011) Elemental formula annotation of polar and lipophilic metabolites using 13C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. *The Plant Journal*, **68**, 364–376.

Golebiewski, W.M. & **Spenser, I.D.** (1988) Biosynthesis of the lupine alkaloids. II. Sparteine and lupanine. *Canadian Journal of Chemistry*, **66**, 1734–1748.

Gong, L., **Chen, W.**, **Gao, Y.**, **Liu, X.**, **Zhang, H.**, **Xu, C.** ***et al.*** (2013) Genetic analysis of the metabolome exemplified using a rice population. *Proceedings of the National Academy of Sciences*, **110**, 20320–20325.

Good, A.G. & **Muench, D.G.** (1993) Long-term anaerobic metabolism in root tissue (metabolic products of pyruvate metabolism). *Plant Physiology*, **101**, 1163–1168.

Guerra-García, A., **Gioia, T.**, **von Wettberg, E.**, **Logozzo, G.**, **Papa, R.**, **Bitocchi, E.** ***et al.*** (2021) Intelligent characterization of lentil genetic resources: evolutionary history, genetic diversity of germplasm, and the need for well-represented collections. *Current Protocols*, **1**, e134.

Hall, R., **Beale, M.**, **Fiehn, O.**, **Hardy, N.**, **Sumner, L.** & **Bino, R.** (2002) Plant metabolomics: the missing Link in functional genomics strategies. *The Plant Cell*, **14**, 1437–1440.

He, J., **Verstappen, F.**, **Jiao, A.**, **Dicke, M.**, **Bouwmeester, H.J.** & **Kappers, I.F.** (2022) Terpene synthases in cucumber (Cucumis sativus) and their contribution to herbivore-induced volatile terpenoid emission. *New Phytologist*, **233**, 862–877.

Heiling, S., **Schuman, M.C.**, **Schoettner, M.**, **Mukerjee, P.**, **Berger, B.**, **Schneider, B.** ***et al.*** (2010) Jasmonate and ppHsystemin regulate key Malonylation steps in the biosynthesis of 17-Hydroxygeranyllinalool Diterpene glycosides, an abundant and effective direct defense against herbivores in Nicotiana attenuata. *The Plant Cell*, **22**, 273–292.

Hemscheidt, T. & **Spenser, I.D.** (1987) Biosynthesis of lupanine: incorporation of [3,3-2H2]cadaverine. *Canadian Journal of Chemistry*, **65**, 170–174.

Hong, W.-J., **Kim, Y.-J.**, **Chandran, A.K.N.** & **Jung, K.-H.** (2019) Infrastructures of systems biology that facilitate functional genomic study in rice. *Rice*, **12**, 15.

Hummel, J., **Segu, S.**, **Li, Y.**, **Irgang, S.**, **Jueppner, J.** & **Giavalisco, P.** (2011) Ultra performance liquid chromatography and high resolution mass spectrometry for the analysis of plant lipids. *Frontiers in Plant Science*, **2**. Available from: https://doi.org/10.3389/fpls.2011.00054

Iqbal, A., **Khalil, I.A.**, **Ateeq, N.** & **Sayyar Khan, M.** (2006) Nutritional quality of important food legumes. *Food Chemistry*, **97**, 331–335.

Jo, H.-J., **Han, J.Y.**, **Hwang, H.-S.** & **Choi, Y.E.** (2017) β-Amyrin synthase (EsBAS) and β-amyrin 28-oxidase (CYP716A244) in oleanane-type triterpene saponin biosynthesis in Eleutherococcus senticosus. *Phytochemistry*, **135**, 53–63.

Khrisanapant, P., **Kebede, B.**, **Leong, S.Y.** & **Oey, I.** (2019) A comprehensive characterisation of volatile and fatty acid profiles of legume seeds. *Foods*, **8**, 651.

Kopka, J., **Schauer, N.**, **Krueger, S.**, **Birkemeyer, C.**, **Usadel, B.**, **Bergmüller, E.** ***et al.*** (2005) GMD@CSB.DB: the Golm metabolome database. *Bioinformatics*, **21**, 1635–1638.

Kroc, M., **Koczyk, G.**, **Kamel, K.A.**, **Czepiel, K.**, **Fedorowicz-Strońska, O.**, **Krajewski, P.** ***et al.*** (2019) Transcriptome-derived investigation of biosynthesis of quinolizidine alkaloids in narrow-leafed lupin (Lupinus angustifolius L.) highlights candidate genes linked to iucundus locus. *Scientific Reports*, **9**, 2231.

Kroc, M., **Tomaszewska, M.**, **Czepiel, K.**, **Bitocchi, E.**, **Oppermann, M.**, **Neumann, K.** ***et al.*** (2021) Towards development, maintenance, and standardized phenotypic characterization of single-seed-descent genetic resources for Lupins. *Current Protocols*, **1**, e191.

Lisec, J., **Schauer, N.**, **Kopka, J.**, **Willmitzer, L.** & **Fernie, A.R.** (2006) Gas chromatography mass spectrometry–based metabolite profiling in plants. *Nature Protocols*, **1**, 387–396.

Ludwig, M., **Nothias, L.-F.**, **Dührkop, K.**, **Koester, I.**, **Fleischauer, M.**, **Hoffmann, M.A.** ***et al.*** (2020) Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nature Machine Intelligence*, **2**, 629–641.

Luo, J. (2015) Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology*, **24**, 31–38.

Mancinotti, D., **Frick, K.M.** & **Geu-Flores, F.** (2022) Biosynthesis of quinolizidine alkaloids in lupins: mechanistic considerations and prospects for pathway elucidation. *Natural Product Reports*, **39**, 1423–1437.

Martin, C. (2018) A role for plant science in underpinning the objective of global nutritional security? *Annals of Botany*, **122**, 541–553.

Martin, C., Butelli, E., Petroni, K. & Tonelli, C. (2011) How can research on plants contribute to promoting human health? *The Plant Cell*, **23**, 1685–1699.

Mugford, S.T. & Osbourn, A. (2013) Saponin synthesis and function. In: Bach, T.J. & Rohmer, M. (Eds.) *Isoprenoid Synthesis in Plants and Microorganisms: New Concepts and Experimental Approaches*. New York, NY: Springer New York, pp. 405–424.

Nakashima, H., Okubo, K., Honda, Y., Tamura, T., Matsuda, S. & Yamamoto, N. (1989) Inhibitory effect of glycosides like saponin from soybean on the infectivity of HIV in vitro. *AIDS*, **3**, 655–658.

Nakayama, T., Takahashi, S. & Waki, T. (2019) Formation of flavonoid metabolons: functional significance of protein-protein interactions and impact on flavonoid chemodiversity. *Frontiers in Plant Science*, **10**. Available from: https://doi.org/10.3389/fpls.2019.00821

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A. *et al.* (2020) Feature-based molecular networking in the GNPS analysis environment. *Nature Methods*, **17**, 905–908.

Oikawa, A., Matsuda, F., Kusano, M., Okazaki, Y. & Saito, K. (2008) Rice metabolomics. *Rice*, **1**, 63–71.

Okazaki, Y. & Saito, K. (2014) Roles of lipids as signaling molecules and mitigators during stress response in plants. *The Plant Journal*, **79**, 584–596.

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

Perez-Fons, L., Bohorquez-Chaux, A., Irigoyen, M.L., Garceau, D.C., Morreel, K., Boerjan, W. *et al.* (2019) A metabolomics characterisation of natural variation in the resistance of cassava to whitefly. *BMC Plant Biology*, **19**, 518.

Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, **11**, 395.

Price, E.J., Drapal, M., Perez-Fons, L., Amah, D., Bhattacharjee, R., Heider, B. *et al.* (2020) Metabolite database for root, tuber, and banana crops to facilitate modern breeding in understudied crops. *The Plant Journal*, **101**, 1258–1268.

Rebello, C.J., Greenway, F.L. & Finley, J.W. (2014) A review of the nutritional value of legumes and their effects on obesity and its related comorbidities. *Obesity Reviews*, **15**, 392–407.

Richter, A., Schaff, C., Zhang, Z., Lipka, A.E., Tian, F., Köllner, T.G. *et al.* (2016) Characterization of biosynthetic pathways for the production of the volatile Homoterpenes DMNT and TMTT in Zea mays. *The Plant Cell*, **28**, 2651–2665.

Rocchetti, L., Bellucci, E., Cortinovis, G., Di Vittori, V., Lanzavecchia, G., Frascarelli, G. *et al.* (2020) The Development of a European and Mediterranean Chickpea Association Panel (EMCAP). *Agronomy*, **10**, 1417.

Rocchetti, L., Gioia, T., Logozzo, G., Brezeanu, C., Pereira, L.G., la Rosa, L.D. *et al.* (2022) Towards the development, maintenance and standardized phenotypic characterization of single-seed-descent genetic resources for chickpea. *Current Protocols*, **2**, e371.

Rupasinghe, H.P.V., Jackson, C.-J.C., Poysa, V., Di Berardo, C., Bewley, J.D. & Jenkinson, J. (2003) Soyasapogenol a and B distribution in soybean (Glycine max L. Merr.) in relation to seed physiology, genetic variability, and growing location. *Journal of Agricultural and Food Chemistry*, **51**, 5888–5894.

Saito, K. & Matsuda, F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annual Review of Plant Biology*, **61**, 463–489.

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T. *et al.* (2013) The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiology and Biochemistry*, **72**, 21–34.

Sakurai, N. (2022) Recent applications of metabolomics in plant breeding. *Breeding Science*, **72**, 56–65.

Salem, M.A., Yoshida, T., Perez de Souza, L., Alseekh, S., Bajdzienko, K., Fernie, A.R. *et al.* (2020) An improved extraction method enables the comprehensive analysis of lipids, proteins, metabolites and phytohormones from a single sample of leaf tissue under water-deficit stress. *The Plant Journal*, **103**, 1614–1632.

Sasaki, K., Mito, K., Ohara, K., Yamamoto, H. & Yazaki, K. (2008) Cloning and characterization of Naringenin 8-Prenyltransferase, a flavonoid-specific Prenyltransferase of Sophora flavescens. *Plant Physiology*, **146**, 1075–1084.

Sawai, S. & Saito, K. (2011) Triterpenoid biosynthesis and engineering in plants. *Frontiers in Plant Science*, **2**. Available from: https://doi.org/10.3389/fpls.2011.00025

Sayama, T., Ono, E., Takagi, K., Takada, Y., Horikawa, M., Nakamoto, Y. *et al.* (2012) The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid Saponins of soybean. *The Plant Cell*, **24**, 2123–2138.

Schäfer, H. & Wink, M. (2009) Medicinally important secondary metabolites in recombinant microorganisms or plants: progress in alkaloid biosynthesis. *Biotechnology Journal*, **4**, 1684–1703.

Schulz, E., Tohge, T., Zuther, E., Fernie, A.R. & Hincha, D.K. (2016) Flavonoids are determinants of freezing tolerance and cold acclimation in Arabidopsis thaliana. *Scientific Reports*, **6**, 34027.

Sebastiana, M., Gargallo-Garriga, A., Sardans, J., Pérez-Trujillo, M., Monteiro, F., Figueiredo, A. *et al.* (2021) Metabolomics and transcriptomics to decipher molecular mechanisms underlying ectomycorrhizal root colonization of an oak tree. *Scientific Reports*, **11**, 8576.

Shen, G., Huhman, D., Lei, Z., Snyder, J., Sumner, L.W. & Dixon, R.A. (2012) Characterization of an Isoflavonoid-specific prenyltransferase from Lupinus albus. *Plant Physiology*, **159**, 70–80.

Shimizu, Y., Rai, A., Okawa, Y., Tomatsu, H., Sato, M., Kera, K. *et al.* (2019) Metabolic diversification of nitrogen-containing metabolites by the expression of a heterologous lysine decarboxylase gene in Arabidopsis. *The Plant Journal*, **100**, 505–521.

Shinbo, Y., Nakamura, Y., Altaf-Ul-Amin, M., Asahi, H., Kurokawa, K., Arita, M. *et al.* (2006) KNApSAcK: a comprehensive species-metabolite relationship database. In: Saito, K., Dixon, R.A. & Willmitzer, L. (Eds.) *Plant metabolomics*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 165–181.

Sumner, L.W., Lei, Z., Nikolau, B.J. & Saito, K. (2015) Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. *Natural Product Reports*, **32**, 212–229.

Sundaramoorthy, J., Park, G.T., Komagamine, K., Tsukamoto, C., Chang, J.H., Lee, J.-D. *et al.* (2019) Biosynthesis of DDMP saponins in soybean is regulated by a distinct UDP-glycosyltransferase. *New Phytologist*, **222**, 261–274.

Sundaramoorthy, J., Park, G.T., Mukaiyama, K., Tsukamoto, C., Chang, J.H., Lee, J.-D. *et al.* (2018) Molecular elucidation of a new allelic variation at the Sg-5 gene associated with the absence of group a saponins in wild soybean. *PLoS One*, **13**, e0192150.

Suzuki, H., Murakoshi, I. & Saito, K. (1994) A novel O-tigloyltransferase for alkaloid biosynthesis in plants. Purification, characterization, and distribution in Lupinus plants. *Journal of Biological Chemistry*, **269**, 15853–15860.

Takagi, K., Nishizawa, K., Hirose, A., Kita, A. & Ishimoto, M. (2011) Manipulation of saponin biosynthesis by RNA interference-mediated silencing of β-amyrin synthase gene expression in soybean. *Plant Cell Reports*, **30**, 1835–1846.

Tanaka, Y., Sasaki, N. & Ohmiya, A. (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal*, **54**, 733–749.

Tohge, T., de Souza, L.P. & Fernie, A.R. (2017) Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *Journal of Experimental Botany*, **68**, 4013–4028.

Tohge, T. & Fernie, A.R. (2015) Metabolomics-inspired insight into developmental, environmental and genetic aspects of tomato fruit chemical composition and quality. *Plant and Cell Physiology*, **56**, 1681–1696.

Tohge, T. & Fernie, R.A. (2017) An overview of compounds derived from the shikimate and phenylpropanoid pathways and their medicinal importance. *Mini-Reviews in Medicinal Chemistry*, **17**, 1013–1027.

Tohge, T., Perez de Souza, L. & Fernie, A.R. (2018) On the natural diversity of phenylacylated-flavonoid and their in planta function under conditions of stress. *Phytochemistry Reviews*, **17**, 279–290.

Tohge, T., Scossa, F., Wendenburg, R., Frasse, P., Balbo, I., Watanabe, M. *et al.* (2020) Exploiting natural variation in tomato to define pathway structure and metabolic regulation of fruit polyphenolics in the lycopersicum complex. *Molecular Plant*, **13**, 1027–1046.

Tohge, T., Watanabe, M., Hoefgen, R. & Fernie, A. (2013) Shikimate and phenylalanine biosynthesis in the green lineage. *Frontiers in Plant Science*, **4**. Available from: https://doi.org/10.3389/fpls.2013.00062

Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R. *et al.* (2016) Characterization of a recently evolved flavonol-

phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nature Communications*, **7**, 12399.

Tripathi, A., **Vázquez-Baeza, Y.**, **Gauglitz, J.M.**, **Wang, M.**, **Dührkop, K.**, **Nothias-Esposito, M.** ***et al.*** (2021) Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nature Chemical Biology*, **17**, 146–151.

Tsai, C.-Y., **Chen, Y.-H.**, **Chien, Y.-W.**, **Huang, W.-H.** & **Lin, S.-H.** (2010) Effect of soy saponin on the growth of human colon cancer cells. *World Journal of Gastroenterology: WJG*, **16**, 3371–3376.

Tsugawa, H., **Ikeda, K.**, **Takahashi, M.**, **Satoh, A.**, **Mori, Y.**, **Uchino, H.** ***et al.*** (2020) A lipidome atlas in MS-DIAL 4. *Nature Biotechnology*, **38**, 1159–1163.

Varshney, R.K., **Thudi, M.**, **Roorkiwal, M.**, **He, W.**, **Upadhyaya, H.D.**, **Yang, W.** ***et al.*** (2019) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nature Genetics*, **51**, 857–864.

Venturelli, S., **Burkard, M.**, **Biendl, M.**, **Lauer, U.M.**, **Frank, J.** & **Busch, C.** (2016) Prenylated chalcones and flavonoids for the prevention and treatment of cancer. *Nutrition*, **32**, 1171–1178.

Vernoud, V., **Lebeigle, L.**, **Munier, J.**, **Marais, J.**, **Sanchez, M.**, **Pertuit, D.** ***et al.*** (2021) β-Amyrin Synthase1 controls the accumulation of the major Saponins present in pea (Pisum sativum). *Plant and Cell Physiology*, **62**, 784–797.

Vincken, J.-P., **Heng, L.**, **de Groot, A.** & **Gruppen, H.** (2007) Saponins, classification and occurrence in the plant kingdom. *Phytochemistry*, **68**, 275–297.

Walle, T. (2007) Methoxylated flavones, a superior cancer chemopreventive flavonoid subclass? *Seminars in Cancer Biology*, **17**, 354–362.

Wang, H., **Xia, C.**, **Chen, L.**, **Zhao, J.**, **Tao, W.**, **Zhang, X.** ***et al.*** (2019) Phytochemical information and biological activities of quinolizidine alkaloids in Sophora: a comprehensive review. *Current Drug Targets*, **20**, 1572–1586.

Wen, W., **Li, D.**, **Li, X.**, **Gao, Y.**, **Li, W.**, **Li, H.** ***et al.*** (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications*, **5**, 3438.

Wink, M. (1983) Wounding-induced increase of quinolizidine alkaloid accumulation in lupin leaves. *Zeitschrift für Naturforschung C*, **38**, 905–909.

Wink, M. (1988) Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. *Theoretical and Applied Genetics*, **75**, 225–233.

Wink, M. (1993) Allelochcmical properties and the raison d'etre ofalkaloids. In: Cordell, G. (Ed.). New York: The alkaloid Academic Press, pp. 1–118.

Wink, M. (2008) Plant secondary metabolism: diversity, function and its evolution. *Natural Product Communications*, **3**, 1934578X0800300801.

Wink, M. (2013) Evolution of secondary metabolites in legumes (Fabaceae). *South African Journal of Botany*, **89**, 164–175.

Wink, M. & **Hartmann, T.** (1982) Enzymatic synthesis of quinolizidine alkaloid esters: a tigloyl-CoA: 13-hydroxylupanine O-tigloyltransferase from Lupinus albus L. *Planta*, **156**, 560–565.

Wink, M. & **Schimmer, O.** (2010) Molecular modes of action of defensive secondary metabolites. *Annual Plant Reviews Volume 39: Functions and Biotechnology of Plant Secondary Metabolites*, **39**, 21–161.

Wink, M. & **Witte, L.** (1985) Quinolizidine alkaloids as nitrogen source for lupin seedlings and cell cultures. *Zeitschrift für Naturforschung C*, **40**, 767–775.

Wymann, M.P. & **Schneiter, R.** (2008) Lipid signalling in disease. *Nature Reviews Molecular Cell Biology*, **9**, 162–176.

Xu, C. & **Shanklin, J.** (2016) Triacylglycerol metabolism, function, and accumulation in plant vegetative tissues. *Annual Review of Plant Biology*, **67**, 179–206.

Yang, T., **Nagy, I.**, **Mancinotti, D.**, **Otterbach, S.L.**, **Andersen, T.B.**, **Motawia, M.S.** ***et al.*** (2017) Transcript profiling of a bitter variety of narrow-leafed lupin to discover alkaloid biosynthetic genes. *Journal of Experimental Botany*, **68**, 5527–5537.

Yano, R., **Takagi, K.**, **Takada, Y.**, **Mukaiyama, K.**, **Tsukamoto, C.**, **Sayama, T.** ***et al.*** (2017) Metabolic switching of astringent and beneficial triterpenoid saponins in soybean is achieved by a loss-of-function mutation in cytochrome P450 72A69. *The Plant Journal*, **89**, 527–539.

Zartl, B., **Silberbauer, K.**, **Loeppert, R.**, **Viernstein, H.**, **Praznik, W.** & **Mueller, M.** (2018) Fermentation of non-digestible raffinose family oligosaccharides and galactomannans by probiotics. *Food & Function*, **9**, 1638–1646.

Zhang, F., **Wu, J.**, **Sade, N.**, **Wu, S.**, **Egbaria, A.**, **Fernie, A.R.** ***et al.*** (2021) Genomic basis underlying the metabolome-mediated drought adaptation of maize. *Genome Biology*, **22**, 260.

Zhou, F. & **Pichersky, E.** (2020) More is better: the diversity of terpene metabolism in plants. *Current Opinion in Plant Biology*, **55**, 1–10.

Zhu, G., **Wang, S.**, **Huang, Z.**, **Zhang, S.**, **Liao, Q.**, **Zhang, C.** ***et al.*** (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261.e212.