Food Research International 209 (2025) 116154

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Investigation of polyphenol diversity among lentil species (*Lens* spp.) using mass spectrometry-based metabolomics guided by photodiode array detection

Randy W. Purves [a,b,c,*], Hamid Khazaei [d,e], Fatma M. Elessawy [a], Roger Munro [c], Bryn O. Shurmer [c], Albert Vandenberg [b]

[a] College of Pharmacy and Nutrition, University of Saskatchewan, 107 Wiggins Road, Saskatoon, SK S7N 5E5, Canada
[b] Department of Plant Sciences, University of Saskatchewan, Agriculture Building 51 Campus Drive Saskatoon SK, S7N 5A8, Canada
[c] Centre for Veterinary Drug Residues, Canadian Food Inspection Agency, 116 Veterinary Road, Saskatoon, SK S7N 2R3, Canada
[d] Production systems, Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790, Helsinki, Finland
[e] Department of Agricultural Sciences, University of Helsinki, Latokartanonkaari 5, 00014, Helsinki, Finland

ARTICLE INFO

Keywords:
Lentils
LC-MS
Untargeted analysis
Polyphenols
Crop wild relatives
Seed quality
Proanthocyanidins

ABSTRACT

Polyphenol diversity was investigated among seven lentil species, including *Lens culinaris* (cultivated lentil), *L. orientalis*, *L. tomentosus*, *L. odemensis*, *L. lamottei*, *L. ervoides*, and *L. nigricans*, using photodiode array detection coupled with liquid chromatography - mass spectrometry (LC-MS). Principal component analysis showed that most species grouped individually, except *L. tomentosus* and *L. odemensis,* which overlapped. The LC-MS data from both negative and positive electrospray ionization modes were used to identify 85 polyphenols observed in the UV–vis spectra, which included 27 proanthocyanidins, 17 flavonols, 15 flavones, and 12 hydroxybenzoic acids. An untargeted (comprehensive) analysis of the LC-MS data using Compound Discoverer software identified additional polyphenols (231 total), including numerous overlapping proanthocyanidins that contribute to a broad peak in the UV–vis spectra. The software analysis uncovered some notable differences among polyphenol profiles and intensities within the flavones, flavonols, and phenolic acids present in the species. This result indicates natural variation among the lentil wild relatives, which in part, is attributed to structurally isomeric compounds. A hierarchical clustering analysis, and a differential analysis using volcano plots used to look for statistically significant differences in polyphenols, illustrated significantly lower relative levels of polyphenols in *L. culinaris* compared with the wild types, especially within the proanthocyanidins and flavones. Our results highlight the potential of lentil wild relatives to enhance lentil seed quality.

1. Introduction

Cultivated lentil (*Lens culinaris* Medik.) is an ancient protein rich and nutritious staple cool-season grain legume crop. Lentil seeds also have high antioxidant activity compared to other grain legume species mainly due to total and specific phenolic compounds (Elessawy et al., 2021; Grela et al., 2017). Lentils are in high demand globally, and have, by far, the highest growth rate in consumption compared to other grain legume crops (Khazaei et al., 2019). According to FAO (FAOSTAT, 2024) in 2022, Canada was the largest producer, contributing over 30 % of the world's production followed by India, Australia, Türkiye, and the United States of America. Lentil breeding has created a genetic bottleneck, particularly for seed quality traits, ultimately restricting breeding output (Khazaei et al., 2016). Cultivated lentils are selectively bred for specific desirable traits like high yield, uniformity, and disease resistance. Crop wild relatives, which represent a vast array of genetic traits, remain underutilized in breeding programs but hold significant potential for overcoming this bottleneck (Bohra et al., 2022). The *Lens* genus has seven closely related taxa (Wong et al., 2015); *L. culinaris* (cultivated lentil), *L. orientalis*, *L. tomentosus* (primary gene pool); *L. odemensis*, *L. lamottei*, (secondary gene pool); *L. ervoides* (tertiary gene pool); and *L. nigricans* (quaternary gene pool). Though many of the wild relatives of lentil have better adaptation to biotic and abiotic stresses (e.g., Barilli & Rubiales, 2023; Bhadauria et al., 2016; Coyne et al., 2020; Podder et al.,

* Corresponding author at: College of Pharmacy and Nutrition, University of Saskatchewan, 107 Wiggins Road, Saskatoon, SK S7N 5E5, Canada.
E-mail address: randy.purves@usask.ca (R.W. Purves).

https://doi.org/10.1016/j.foodres.2025.116154
Received 25 November 2024; Received in revised form 31 January 2025; Accepted 7 March 2025
Available online 13 March 2025

2013; Singh et al., 2020) and higher nutritional value (e.g., Kumar et al., 2018; Zhang et al., 2019) than cultivated lentils, they have been neglected and underutilized.

Polyphenols represent a large class of compounds in lentils that warrant particular interest as an important group of plant secondary metabolites that act as defense mechanisms against environmental conditions such as UV radiation, pathogens, and pests (Dixon, 2001; Farkas & Kiraaly, 1962; Nicholson & Hammerschmidt, 1992). Polyphenols are derived from phenylalanine or shikimic acid, typically are conjugated, and may contain one or more sugar residues (Pandey & Rizvi, 2009). Polyphenols are classified into different groups as a function of the number of phenol rings and on the basis of structural elements that bind these rings (Pandey & Rizvi, 2009). Polyphenols have been found to be particularly abundant in lentil seed coat (Dueñas et al., 2002). Common polyphenol classes found in lentils include the phenolic acids and flavonoids, with the latter being the most abundant (Dueñas et al., 2002; Elessawy et al., 2021). Flavonoids, whose structure contains two aromatic rings linked by a 3-carbon bridge (C6-C3-C6) (Mustafa et al., 2022), can be further subdivided, and categorized into monomers or oligomers. Flavonols are typically the most abundant monomers (Mustafa et al., 2022), whereas proanthocyanidins (oligomers of flavan-3-ols) are the most abundant oligomers (Elessawy et al., 2021).

Polyphenol profiles of cultivated lentil seed coats have been characterized using a variety of liquid chromatography – mass spectrometry (LC-MS) methods, starting with Dueñas et al. (2002) who used photodiode array (PDA) detection along with LC-MS to identify major polyphenols in the lentil cotyledon and seed coat (Dueñas et al., 2002). Major polyphenols are often quantified using targeted methods, which require authentic standards, and many studies have also examined the relationship between antioxidant properties and these profiles (e.g., Aguilera et al. (2010)). Our lab has employed targeted LC-MS methods (Elessawy et al., 2020; Mirali et al., 2014) using multiple reaction monitoring (MRM) to investigate polyphenols and certain pathways related to the seed coat genetics (Mirali et al., 2016; Mirali et al., 2017). However, more recently, we have also implemented accurate mass instrumentation in an untargeted LC-MS workflow to more comprehensively identify polyphenols. Accurate mass instrumentation (Rathahao-Paris et al., 2016) enables the precise determination of mass-to-charge ratio (m/z), which simplifies the determination of the chemical formula and structure of unknown compounds through tandem mass spectrometry (MS/MS). Although accurate mass instrumentation has been used to identify polyphenols in lentils (Elessawy et al., 2021; Farag et al., 2019; Guo et al., 2022; Llorach et al., 2019; Żuchowski et al., 2014), the usage of untargeted workflows has recently accelerated with the rapid advancement of powerful software that has allowed for more in-depth analysis. Llorach et al. (2019) used an untargeted LC-MS analysis to compare phytochemicals among three legumes. Our group used untargeted LC-MS to compare polyphenols in the seed coats among five pulse crops (Elessawy et al., 2021), and to compare among patterned seed coats in lentils, which showed a high abundance of flavones when a pattern was present (Elessawy, Wright, et al., 2023). Metabolites identified/putatively identified in our previous work have been added to an in-house library to simplify metabolite identification in subsequent studies.

Thus, although many studies have investigated polyphenols in lentil seeds, they have mainly focused on cultivated lentil germplasm. To our knowledge, this is the first in-depth study probing polyphenol differences among all wild lentil species and comparing them with the cultivated form. To thoroughly probe these differences, a combination of approaches was used. A targeted semi-quantitative LC-MS method was used (Elessawy et al., 2020), along with an untargeted LC-MS analysis, which was guided by PDA detection. The main goals of this study were to identify the most abundant polyphenols among all seven lentil species seeds and to look for key differences among them. Understanding these differences could guide lentil improvement programs to enhance lentil nutrition and health benefits.

2. Materials and methods

2.1. Plant material and growing conditions

The plant materials consisted of 24 genotypes of *L. culinaris* (cultivated lentil), six genotypes from each of *L. ervoides*, *L. orientalis*, *L. tomentosus, L. odemensis*, and *L. nigricans,* and five genotypes of *L. lamottei*. The 59 lentil genotypes investigated in this study are shown in **Supplementary Table S1**. Lentil seeds used in this study were harvested from plants grown in a climate-controlled phytotron (https://agbio.usask.ca/research/centres-and-facilities/controlled-environment-facility.php) in the College of Agriculture and Bioresources controlled environment facility at the University of Saskatchewan, Canada in 2018. The photoperiod was adjusted to 16 h day and 8 h night, with the temperature set to 21 °C (day) and 16 °C (night) to simulate optimal growing conditions for lentils, which generally thrive in cooler climates. The photon flux was set to 300 μmol m^{-2} s^{-1} during the light period. Plants were grown in 10-cm pots filled with growing mixture # 3 (Sun-Gro Horticulture, Agawam, MA, USA). Soil moisture level was maintained at field capacity with automatic irrigation for all plants. These growing conditions were chosen to simulate optimal growing conditions for lentils, which generally thrive in cooler climates.

2.2. Chemicals and reagents

Standards and isotopically labelled internal standards for the semi-quantitative targeted LC-MS method were sourced and used as described previously (Elessawy et al., 2020). Optima LC-MS grade acetone, acetonitrile, and methanol along with Acros Organics formic acid were obtained from Fisher Scientific (Nepean, ON, Canada). Water was obtained from a Millipore Milli-Q water system (Milford, MA, USA).

2.3. Seed extraction and preparation of samples

Approximately 100 mg of crushed whole seeds were weighed to ±1 mg in 2 mL Eppendorf tubes. The seeds were pulverized and extracted using a procedure described previously for our targeted method (Elessawy et al., 2020), which uses six isotopically labelled internal standards for quantification; namely quercetin-D_3, ± − catechin-2,3,4-$^{13}C_3$, resveratrol-(4-hydroxyphenyl-$^{13}C_6$, vanillin-(ring-$^{13}C_6$), 4-hydroxybenzoic acid-$^{13}C_7$, and ferulic acid-D_3. After centrifugation, 200 μL of supernatant was placed into a separate Eppendorf tube that was immediately dried down and stored at −80 °C until analysis to minimize degradation. On the day of analysis (next day in this study), dried-down supernatant was reconstituted in 200 μL of 9:1 H_2O:MeOH (Mirali et al., 2014) and transferred to an HPLC vial. Immediately after the targeted analysis was complete, the untargeted analysis (described below) was carried out used the same extracts.

A quality control (QC) sample for the untargeted analysis was prepared by taking 10 μL of extract from each genotype, adding to a single tube (590 μL total), and vortexing for 15 s to mix thoroughly. The QC was dried down and reconstituted in 590 μL of 9:1 H_2O:MeOH and transferred to an HPLC vial on the day of analysis. The implementation of a QC sample has been described previously (Dunn et al., 2011). In brief, the QC contains all of the unknown compounds and therefore can be used for relative quantification of the unknowns as described in Section 2.4.2. In addition, seven "Identification only" (ID) samples were prepared by adding 10 μL (*L. culinaris*) or 30 μL (other *Lens* species) of each genotype to a tube specific for that wild lentil species (i.e., one tube for each species). These were also dried down and reconstituted individually in 9:1 H_2O:MeOH.

2.4. LC-MS methods

Thermo Fisher's XCalibur 4.1 software was used to acquire and view data for both LC-MS methods.

2.4.1. Targeted LC-MS

The targeted LC-MS method used for quantification has been described previously (Elessawy et al., 2020). This method was developed for use with pulse crops based on polyphenols reported in the literature. In short, an Agilent (Santa Clara, CA, USA) 1290 UHPLC was coupled to a Thermo Fisher (San Jose, CA, USA) TSQ Vantage triple quadrupole mass spectrometer. An Agilent poroshell 120 PFP (2.1 mm × 100 mm, 2.7 μm) column was used based on separation capabilities and peak shapes as described previously (Elessawy et al., 2020). The 30-min gradient is shown in **Supplementary Table S2**. MS detection was achieved using multiple reaction monitoring (MRM) for 98 polyphenols from 13 different classes. The complete method, including transitions and instrument conditions is described elsewhere (Elessawy et al., 2020).

2.4.2. Untargeted LC-MS guided by PDA

To identify all major polyphenols present in the lentil species, an untargeted LC-MS method was used, which followed a similar procedure to our previous work (Elessawy et al., 2021; Elessawy, Wright, et al., 2023). The untargeted instrumentation available in our lab uses a Thermo Fisher Ultimate 3000 UHPLC coupled to a Thermo Fisher Q-Exactive high resolution mass spectrometer (HRMS). The Ultimate 3000 UHPLC is equipped with a PDA detector, and since all polyphenols absorb UV–vis radiation in the region of 250–600 nm, the PDA was used to guide the LC-MS analysis as was described previously (Elessawy, Hughes, et al., 2023; Zanotto et al., 2020; Zanotto et al., 2023).

The untargeted analysis uses the same LC column and conditions as were used for the targeted analysis. The Q-Exactive was operated in both positive and negative modes (in separate injections) since some compounds might only ionize in one mode. Each of the 59 genotypes were analyzed using full scan MS (*m/z* 140–2100) at the highest mass resolution available (140,000 resolution at m/z 200) for accurate mass measurements. The experiments were set up such that the QC sample was injected every 9th sample. Thus, the QC can account for differences in instrument response and enable relative quantification as described elsewhere (Dunn et al., 2011).

The seven ID samples were used to obtain fragmentation data, which was used to identify compounds by comparing with MS/MS libraries. These data dependent (DD) experiments were carried out using the scan function "Full scan/DDMS2". A full scan mass spectrum (70,000 resolution @ m/z 200) identified the seven most intense ions for MS/MS analysis, then the corresponding MS/MS analyses used stepped normalized collision energy (NCE) values of 30 and 50 eV (17,500 resolution at m/z 200). The MS/MS analysis used a dynamic exclusion time of 3 s (about half the width of the narrowest LC peak). Dynamic exclusion helps to prevent the acquisition of repetitive MS/MS spectra and allows for fragmentation of more compounds. The ID samples were placed at the beginning and the end of the sample queue. Intense background ions in the blank were used to create an exclusion list in the first set of ID samples. After the ID samples were acquired at the start of the queue, these samples were analyzed by the Compound Discoverer software (described below) while the queue was running to determine compounds that had at least two MS/MS spectra. The *m/z* values (and their isotopes) for compounds having at least two MS/MS spectra were added to the exclusion list for the second acquisition of the ID samples at the end of the queue.

2.5. Data analysis

2.5.1. Targeted LC-MS

Calculations of absolute polyphenol amounts were done using the quantification software QuanBrowser available within Thermo XCalibur 4.1, as was described previously (Elessawy et al., 2020). Only polyphenols that were above the lower limit of quantification in at least one of the lentil species were reported.

2.5.2. Identification using untargeted LC-MS & PDA

A customized untargeted data analysis workflow adapted from the Thermo Fisher Compound Discoverer 3.3 software (CD 3.3) was used to process the raw data. We have reported the use of a similar workflow previously (Elessawy et al., 2021; Elessawy, Wright, et al., 2023); the main difference in this work is the use of a newer version of the software. The CD 3.3 parameters used in this study are summarized by the software in text format and this information is summarized in the Supporting Information as **Supplementary Summary S1**.

The CD 3.3 workflow determines possible molecular formulae for each m/z value (feature) using the full-scan accurate mass data. Since features are dependent on the ionization mode, negative and positive data are analyzed separately. MS/MS spectra from the ID samples are used for identification purposes. In addition to using Thermo Fisher's mzCloud library, which contains fragmentation data of over 32,000 compounds analyzed with Thermo Orbitrap instrumentation (www.mzcloud.org), MS/MS spectra from our in-house libraries were also compared (using the mzVault node). When necessary, a manual comparison of MS/MS data with on-line compound libraries (e.g., FoodB) was also carried out.

The PDA and LC-MS data were manually aligned to identify the major polyphenols. Polyphenols were identified with varying levels of confidence, by comparing their MS/MS spectra, acquired at low (10/20 eV) and high (30/50 eV) collision energies, with online and in-house databases. In some instances, additional higher energy spectra (50/70 or 70/90 eV) were needed for identification and acquired after the original data acquisition.

2.5.3. More extensive identification of polyphenols in NEG mode

A CD 3.3 analysis generated a table of thousands of compounds. After identifying the 85 most abundant polyphenols that corresponded to the UV peaks, we wanted to include additional polyphenols in a hierarchical analysis. Thus, as was done previously, the CD 3.3 results were used with filters to restrict the analysis of compounds to a workable number; filters included polyphenols having values ≥ 50 million counts for maximum area, and those that eluted between 2 and 18 min (Elessawy et al., 2021). With these filters, the compound table generated by the software showed some entries representing adduct peaks, fragment peaks, and isotope peaks that did not group (as desired) with the corresponding molecular ion. After manually removing these entries, 176 compounds were identified as polyphenols.

In addition, because of the predictability of *m/z* values and MS/MS spectra for proanthocyanidin oligomers (namely procyanidins and prodelphinidins as described in Section 3.2), a lower maximum area value (i.e., ≥ 5 million counts) was used for detecting proanthocyanidins. A total of 111 proanthocyanidins were identified above this cut-off area. Finally, the CD software results were compared with the polyphenols identified with the targeted method. Of the 29 polyphenols listed in **Supplementary Table S3**, 26 were identified with NEG mode using the software. The result of combining the three approaches was that a total of 231 polyphenols were identified (some polyphenols were identified in more than one approach) and used for the hierarchical clustering analysis.

2.5.4. Statistical analysis

Principal Component Analysis (PCA) plots were generated by CD 3.3. The scores plot and loadings plot both used normalized areas and plots of PC1 vs. PC2.

A Hierarchical Clustering Analysis (HCA) was used in a similar fashion as described previously (Elessawy, Hughes, et al., 2023). In brief, HCA uses an agglomerative (bottom-up) approach to find the similarities between the samples and compounds. We used normalized areas and "scale before clustering", which applies a z-score transformation before performing the hierarchical clustering. With this scaling, the heat map legend displays the range of the scaled values, and the dendrogram nodes display the scaled distance values. The remaining

user-defined variables in the HCA analysis for "Distance function" and "Linkage method" were the defaults "Euclidean" and "Complete", respectively.

A volcano plot is a type of scatterplot that visually shows differential analysis when comparing between two groups. Statistical significance is plotted on the y-axis and magnitude of change is plotted on the x-axis. In this study, for a polyphenol to be considered significantly different, the relative peak areas (calculated for each replicate within a group and the median value was used) need to be $\geq$ 4.0 times ($\log_2$ fold change = 2) different and the *P*-value < 0.001 (>99.9 % confidence). *P*-values per group ratio were calculated by ANOVA and TukeyHSD post hoc tests.

3. Results and discussion

3.1. Exploring polyphenol diversity using LC-MS with PDA detection

Two LC-PDA-HRMS analyses (one using positive and one negative ionization) were used to investigate each of the 59 lentil genotypes spanning the seven lentil species (ID samples). **Supplementary Fig. S1** shows A) positive and B) negative LC-MS spectra for the *L. culinaris* ID sample. Although these mass spectra are similar, there are some important differences, which will be discussed in more detail when discussing polyphenol classes. An in-line PDA also acquired UV–vis spectra for these analyses, and **Supplementary Fig. S1** (bottom panel) shows the corresponding UV–vis spectrum for the *L. culinaris* ID sample. Although negative mode gave a more complete picture of the polyphenol content compared with positive mode, as will be discussed in more detail in the next section, the use of *both* modes gave the most complete identification. **Supplementary Fig. S1** illustrates that several intense peaks in the MS spectrum are not detected by the PDA and therefore these compounds are not polyphenols since *all* polyphenols absorb between 250 and 600 nm. Note that there is a small offset between the PDA and MS spectra (~0.05 min) due to the volume of tubing that connects the outlet of the PDA to the ionization source. For clarity, retention times reported are from the PDA. Notably, peaks near the void volume (~1 min) or near the column flush/column re-equilibration (25–28 min) are intense in the MS spectra, but significantly less with the PDA. Intense peaks in the void volume were observed at *m/z* 341, 503, 665, and 827 (and their formic acid adducts, all not shown). These peaks are consistent with the presence of disaccharides and oligosaccharides, namely sucrose, raffinose, stachyose, and verbacose, respectively, which are abundantly present and have been characterized previously (Gulewicz et al., 2000; Quemener & Brillouet, 1983; Tsopmo & Muir, 2010). In addition to polyphenols, saponins and phytosterols are also abundant in lentils and these less polar compounds typically elute later in the LC-MS analysis. Two large late eluting peaks in **Supplementary Fig. S1** (middle panel) are saponins; soyasaponin I elutes at Rt = 21.7 min and soyasaponin VI at Rt = 23.4 min. Many (but not all) of the saponin and phytosterol peaks present in the mass spectra do not have corresponding UV–vis peaks (such as soyasaponin I) and are therefore readily eliminated as possible polyphenols.

3.2. Compound identification using CD 3.3 guided by the PDA data

Separate CD 3.3 analyses were set up using positive and negative LC-MS data as is described in Section 2.5. Each analysis grouped features

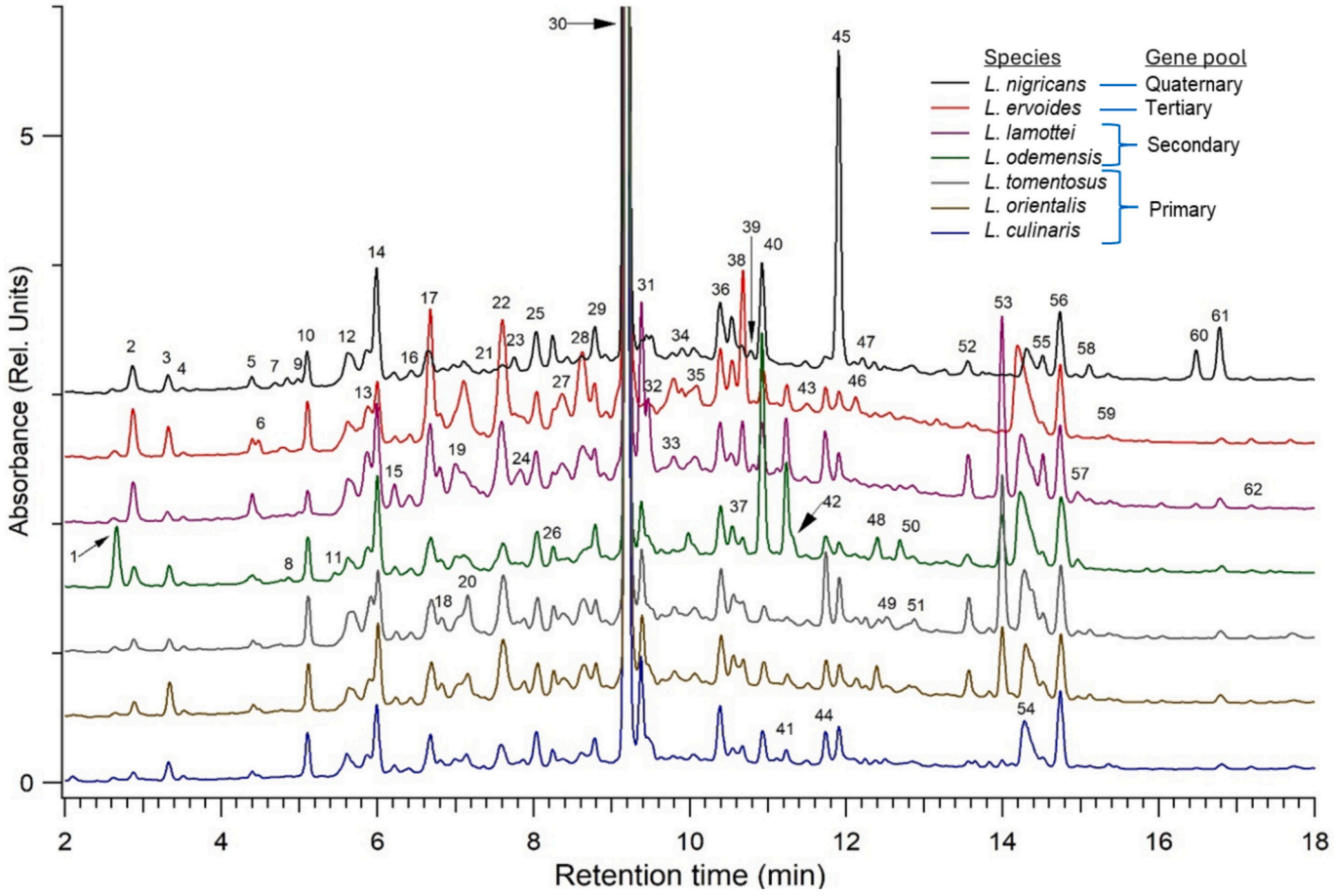

Fig. 1. LC-PDA UV–vis spectra for the seven lentil species using the seven ID samples. The major peaks are numbered and their identities, including ID level, are given in Table 1. The inset shows the species were ordered starting with *L. culinaris* at the bottom. Moving up the traces, the species become more distantly related as indicated by their gene pools (which were derived from Wong et al., 2015). Each subsequent trace is offset by 0.5 absorbance units.

and determined the likely chemical formula for a compound based on accurate mass measurements; compound identities were based on matching fragment ions to on-line libraries and databases (e.g., mzCloud and mzVault). Initially, the PDA was used to focus the analysis on the most intense polyphenols by restricting the MS analysis to regions in which a UV–vis peak was observed. Fig. 1 shows LC-PDA spectra for the ID samples of the seven lentil species (*Lens* spp.). Since the UV peak at 9.2 min dominates the spectrum (**Supplementary Fig. S1**- bottom panel), Fig. 1 shows a 10× expansion of the baseline over the retention range of 2–18 min, which is where the majority of the polyphenols elute using this method (Elessawy et al., 2021). As there are 59 individual genotypes, for clarity, the ID samples for each of the seven species were chosen for presenting the data. As described in Section 2.3, the ID samples contain equal amounts of every genotype in a species, and thus, these traces approximate an *average* spectrum for each species. The ID traces are arranged such that in moving from the bottom trace to the top, the species become increasingly more distantly related to *L. culinaris* based on likely gene pools for the genus *Lens* species using genotyping by sequencing (Wong et al., 2015). At each of the retention times for the peaks numbered in Fig. 1, the LC-PDA traces were compared with compound(s) identified at the same retention time(s) in CD 3.3 using separate positive and negative mode analyses. The CD results for each peak of interest were manually checked and the results are summarized in Table 1, which presents the identities of the major polyphenol(s) corresponding to each of these peak numbers and a corresponding ID level (discussed further below). When more than 1 major polyphenol contributed to a peak in Fig. 1, the second and third polyphenols (where applicable) were labelled as b and c (e.g., 28, 28b, 28c). Table 1 shows 91 polyphenols; 85 which were present in the samples plus the 6 internal standards that were added for the targeted analysis (as described in Section 2.3). All 85 polyphenols were present in the negative mode CD analysis, although tryptophan was at a much lower abundance compared with positive mode. In addition, malonylglycosides typically showed significant fragmentation (loss of CO_2) and were incorrectly identified as acetylglycosides. Conversely, in positive mode, the hydroxybenzoic acids were significantly less intense. Thus, in the positive mode CD analysis, out of the ten acids reported in Table 1 (not including the ISs), three showed up only as sodium adducts, and five were not detected at all. In general, adduct formation and in-source fragmentation were more of an issue in positive mode, although malonylglycosides showed little to no fragmentation, and were correctly identified. Thus, incorporating results from both positive and negative analyses gave more complete results compared with either mode alone.

The identification levels in Table 1 follow Sumner et al. (2007) with an additional level (2/3) that was described in our previous work (Elessawy, Wright, et al., 2023). Compound identities that were confirmed with authentic standards have been assigned as level 1 in Table 1. Note that six isotopically labelled polyphenols, which were added to the extraction solvent as internal standards for quantification in the targeted analysis, are labelled as IS in Table 1, but were excluded in the statistical analysis (Section 3.3). Level 2 represents putatively identified compounds (references are given in Table 1 for compounds putatively identified based on prior work), level 3 represents identification of class only, which indicates the class but not the structure (i.e., hydroxybenzoic acid derivative), and level 4 are unidentified compounds. Level 2/3 was added for two instances in which identification confidence falls between levels 2 and 3. The first use of level 2/3 is for a glycosylated metabolite with a putative aglycone identification, but with only a partial identification (e.g., hexose, pentose) of the glycosylation. The second use of level 2/3 is for isomeric structures, such as for the proanthocyanidins, where without standards, isomeric monomers can not be differentiated, as described further below.

Fig. 2 illustrates examples for the uses of a level 2/3 identification in this study, since many compounds in Table 1 were identified at this level. The left-hand side of the figure shows an example for a glycosylated polyphenol. Fig. 2A shows an MS/MS spectrum for peak #57 (Table 1) that was acquired using a low collision energy (10 eV), which is useful in identifying neutral losses due to glycosylation. The neutral loss of 162.0526 in the figure is within experimental error for the expected neutral loss for a hexose of 162.0528. However, from the MS/MS spectra, neither the exact type of glycosylation nor the location of attachment to the aglycone can be determined; it is only known that a hexoside is present. A higher collision energy (30/50 eV) spectrum in Fig. 2B is dominated by the aglycone peak with some low intensity fragment ions (shown in a red rectangle). An even higher energy fragmentation spectra (50/60 eV) provides a fingerprint of the aglycone as is shown in Fig. 2C. Ions in the MS/MS fingerprint of the aglycone are used to distinguish among structural isomers, including *m/z* 133.0283 (circled in red), which distinguishes luteolin from kaempferol. The aglycone in this compound is putatively identified as luteolin. The fragmentation of the compound is extremely similar to luteolin 4'-O-glucoside (peak 55, for which we have an authentic standard), but elutes at a different retention time, consequently it is assigned as luteolin hexoside and assigned a 2/3 identification.

Proanthocyanidins that could not be confirmed with standards were also identified at level 2/3 as is illustrated on the right-hand side of Fig. 2. Proanthocyanidins are highly abundant in lentils and the two main types are procyanidins and prodelphinidins. Procyanidins are oligomers of catechin and epi-catechin monomers whereas prodelphinidins also contain at least one gallocatechin or epi-gallocatechin. In describing proanthocyanidins in this study, as has been done previously (Elessawy et al., 2021), for simplicity, the notation G indicates that either epi-gallocatechin or gallocatechin is present, whereas C refers to a catechin or epi-catechin. For the compound names given in Table 1, for simplicity, G is placed first. Thus, a GGC trimer indicates the presence of two total gallocatechins/epi-gallocatechins and one catechin/epi-catechin, but how these are arranged is unknown. Since proanthocyanidin m/z values are predictable, they can be readily identified using the full scan and MS/MS spectra. Fig. 2D shows an MS/MS spectrum for procyanidin B3, which is a catechin dimer, catechin-(4α → 8)-catechin (Fig. 1, peak 20). This unknown polyphenol (also compound #20 in Table 1) was identified at level 1 using an authentic standard. The MS/MS fingerprint in Fig. 2D shows several characteristic ions for procyanidin oligomers (e.g., 125.0232, 161.0235, 289.0720, and 407.0778 are most abundant). Although we were unable to source any authentic prodelphinidin standards, we were able to identify prodelphinidin oligomers at a 2/3 level because of the predictability of their m/z values combined with the similarity of their MS/MS spectra with procyanidins. Fig. 2E shows an MS/MS spectrum for GC (corresponding to peak #13b in Table 1), and many of the characteristic ions in the MS/MS spectrum for procyanidin B3 are present (common fragment ions have asterisks). A couple of additional ions, in particular at 177.0185, are also abundant due to the presence of G. Even for the larger oligomers, such as the prodelphinidin hexamer (peak #24b in Table 1), the fingerprint ions are still very similar as is shown in Fig. 2F. Based on the m/z and fragmentation pattern, this compound (peak #24b) is identified at level 2/3 as GGGCCC. Table 1 will be discussed in more detail when examining compound classes in Section 3.4.

3.3. Comparing wild lentil species using absolute and relative quantification

In addition to the untargeted analysis, a targeted method (Elessawy et al., 2020) was used to analyze the same lentil extracts and identified 29 polyphenols that were detected in at least one of the lentil species. The average values for these polyphenols for each of the species are reported in **Supplementary Table S3** along with their molecular weight, chemical formula, and retention time. Importantly, only 12 out of the 85 polyphenols detected using LC-PDA-MS were identified using the targeted method (indicated by Peak# in **Supplementary Table S3**). Since most of the 85 polyphenols were not detected, this illustrates a main limitation of using a targeted method for this type of analysis (i.e.,

Table 1
Identification of polyphenols, sorted by class, from the PDA spectra shown in Fig. 1 using LC-PDA-HRMS. The peak number and retention time are matched with the PDA spectra; the second and third polyphenols (where applicable) are labelled as b and c (e.g., 28, 28b, 28c). The calculated monoisotopic mass, of the uncharged polyphenol is determined based on the observed *m/z* value. The chemical formula is determined based on the accurate mass and the corresponding error is given in ppm. Identification levels are as follows: level 1 is confirmed with an authentic standard, level 2 is putative (based on the fragmentation pattern and literature), level 2/3 is isomeric, level 3 is class only, and level 4 is unidentified. References are given for putative identifications.

Peak	Retention Time (min)	Calculated Monoisotopic Mass	Chemical Formula	Error (ppm)	Identity	ID Level	Reference
Procyanidins							
5b	4.4	866.2086	$C_{30}H_{26}O_{12}$	3.26	Procyanidin trimer (CCC)	2/3	
20	7.1	578.1428	$C_{30}H_{26}O_{12}$	0.70	Procyanidin B3	1	
22	7.55	866.2073	$C_{45}H_{38}O_{18}$	1.69	Procyanidin trimer (CCC)	2/3	
28	8.6	578.1433	$C_{30}H_{26}O_{12}$	1.50	Procyanidin dimer (CC)	2/3	
28b	8.6	1154.2714	$C_{60}H_{50}O_{24}$	1.88	Procyanidin tetramer (CCCC)	2/3	
33	9.8	1154.2712	$C_{60}H_{50}O_{24}$	1.75	Procyanidin tetramer (CCCC)	2/3	
35	10.05	1730.4003	$C_{90}H_{74}O_{36}$	3.41	Procyanidin hexamer (CCCCCC)	2/3	
Prodelphinidins							
8	4.85	594.1387	$C_{30}H_{26}O_{13}$	2.24	Prodelphinidin dimer (GC)	2/3	
8b	4.85	610.1337	$C_{30}H_{26}O_{14}$	2.37	Prodelphinidin dimer (GG)	2/3	
9b	5.0	914.1935	$C_{45}H_{38}O_{21}$	3.23	Prodelphinidin trimer (GGG)	2/3	
13	5.9	898.1975	$C_{45}H_{38}O_{20}$	2.11	Prodelphinidin trimer (GGC)	2/3	
13b	5.9	594.1378	$C_{30}H_{26}O_{13}$	0.72	Prodelphinidin dimer (GC)	2/3	
13c	5.9	756.1911	$C_{36}H_{36}O_{18}$	1.23	Prodelphinidin dimer (GC) hexoside	2/3	
15	6.2	1202.2568	$C_{60}H_{50}O_{27}$	2.40	Prodelphinidin tetramer (GGGC)	2/3	
16	6.45	1506.3147	$C_{75}H_{62}O_{34}$	1.03	Prodelphinidin pentamer (GGGGC)	2/3	
16b	6.45	882.2026	$C_{45}H_{38}O_{19}$	2.16	Prodelphinidin trimer (GCC)	2/3	
17	6.65	882.2021	$C_{45}H_{38}O_{19}$	1.54	Prodelphinidin trimer (GCC)	2/3	
18	6.8	1186.2610	$C_{60}H_{50}O_{26}$	1.66	Prodelphinidin tetramer (GGCC)	2/3	
19	7.0	594.1380	$C_{30}H_{26}O_{13}$	1.17	Prodelphinidin dimer (GC)	2/3	
19b	7.0	1490.3199	$C_{75}H_{62}O_{33}$	1.72	Prodelphinidin pentamer (GGGCC)	2/3	
24	7.8	882.2027	$C_{45}H_{38}O_{19}$	2.21	Prodelphinidin trimer (GCC)	2/3	
24b	7.8	1778.3854	$C_{90}H_{74}O_{39}$	2.61	Prodelphinidin hexamer (GGGCCC)	2/3	
27	8.35	1458.3310	$C_{75}H_{62}O_{31}$	2.40	Prodelphinidin pentamer (GCCCC)	2/3	
27b	8.35	1762.3897	$C_{90}H_{74}O_{38}$	2.19	Prodelphinidin hexamer (GGCCCC)	2/3	
32	9.45	1458.3306	$C_{75}H_{62}O_{31}$	2.15	Prodelphinidin pentamer (GCCCC)	2/3	
33b	9.8	2034.4570	$C_{105}H_{86}O_{43}$	1.36	Prodelphinidin heptamer (GCCCCCC)	2/3	
Flavan-3-ols							
11b	5.50	306.0746	$C_{15}H_{14}O_7$	2.01	(−)-Gallocatechin	1	
17b	6.65	452.1321	$C_{21}H_{24}O_{11}$	1.54	Catechin 3-*O*-glucoside	2	(Zou et al., 2011)
21	7.35	290.0795	$C_{15}H_{14}O_6$	1.52	(+)-Catechin	1	
21b	7.35	293.0895	$^{13}C_3C_{12}H_{14}O_6$	1.50	(±)-Catechin-2,3,4-$^{13}C_3$	1	IS
Hydroxybenzoic acids							
2	2.90	300.0850	$C_{13}H_{16}O_8$	1.54	Hydroxybenzoic acid hexoside	2/3	
5	4.40	330.0957	$C_{14}H_{18}O_9$	1.86	Vanillic acid 4-*O*-glucoside	1	
6	4.50	432.1275	$C_{18}H_{24}O_{12}$	1.61	Hydroxybenzoic acid pentoside hexoside	2/3	
7	4.70	462.1381	$C_{19}H_{26}O_{13}$	1.55	Hydroxybenzoic acid di-hexoside	2/3	
9	5.00	300.0849	$C_{13}H_{16}O_8$	1.10	Hydroxybenzoic acid hexoside	2/3	
12	5.60	145.0529	$^{13}C_7H_6O_3$	0.67	$^{13}C_7$–4-hydroxybenzoic acid	1	IS
12c	5.65	478.1326	$C_{22}H_{18}O_9$	0.78	Dihydroxybenzoic acid di-hexoside	2/3	
20b	7.15	448.1219	$C_{18}H_{24}O_{13}$	0.50	Dihydroxybenzoic acid pentoside hexoside	2/3	
23	7.70	286.0692	$C_{12}H_{14}O_8$	1.34	Gentisic acid 5-*O*-apioside	2	(Tsopmo & Muir, 2010)
25	8.05	158.0677	$^{13}C_6C_2H_8O_3$	1.27	Vanillin-(ring-$^{13}C_6$)	1	IS
29	8.80	418.1115	$C_{17}H_{22}O_{12}$	0.79	Gentisic acid 5-*O*-[β-apiofuranosyl-(1→2)-*O*-β-xylopyranoside]	2	(Tsopmo & Muir, 2010)
57b	14.95	642.2175	$C_{29}H_{38}O_{16}$	2.33	Dihydroxybenzoic acid derivative	3	
Flavonols							
24c	7.85	786.1880	$C_{33}H_{38}O_{22}$	3.25	Kaempferol 3-*O*-β-D-glucopyranosyl(1→2)-β-D-galactopyranoside-7-*O*-β-D-glucuropyranoside	2	(Żuchowski et al., 2014)
30	9.20	902.2698	$C_{39}H_{50}O_{24}$	0.70	Kaempferol di-deoxyhexoside di-hexoside	2	(Taylor et al., 2007)ŋ
31	9.35	756.2124	$C_{33}H_{40}O_{20}$	1.42	Kaempferol deoxyhexoside di-hexoside	2/3	(Tsopmo & Muir, 2010)†
34b	9.95	772.2078	$C_{33}H_{40}O_{21}$	2.10	Kaempferol tri-hexoside	2/3	
38	10.70	756.2128	$C_{33}H_{40}O_{20}$	2.03	Kaempferol deoxyhexoside di-hexoside	2/3	(Tsopmo & Muir, 2010)†
39	10.80	1064.3041	$C_{48}H_{56}O_{27}$	2.99	Kaempferol caffeoyl di-deoxyhexoside di-hexoside	2/3	
39b	10.80	626.1496	$C_{27}H_{30}O_{17}$	2.07	Quercetin di-hexoside	2/3	
40	10.90	740.2176	$C_{33}H_{40}O_{19}$	1.63	Kaempferol 3-*O*-robinoside-7-*O*-rhamnoside (robinin)	1	
41	11.20	740.2178	$C_{33}H_{40}O_{19}$	1.92	Robinin isomer	2/3	

(*continued on next page*)

Table 1 (*continued*)

Peak	Retention Time (min)	Calculated Monoisotopic Mass	Chemical Formula	Error (ppm)	Identity	ID Level	Reference
44	11.75	1048.3088	$C_{48}H_{56}O_{26}$	2.76	Kaempferol coumaroyl di-deoxyhexoside di-hexoside	2/3	(Żuchowski et al., 2014)§
44b	11.75	610.1550	$C_{27}H_{30}O_{16}$	2.67	Quercetin 3-*O*-rutinoside (rutin)	1	
45	11.90	1048.3081	$C_{48}H_{56}O_{26}$	2.01	Kaempferol coumaroyl di-deoxyhexoside di-hexoside	2/3	(Żuchowski et al., 2014)§
47	12.20	1274.3944	$C_{59}H_{70}O_{31}$	3.40	Kaempferol derivative	3	
47b	12.25	1126.3783	$C_{51}H_{66}O_{28}$	3.79	Kaempferol derivative	3	
48	12.40	578.1651	$C_{27}H_{30}O_{14}$	2.61	Kaempferol 3,7-*O*-di-rhamnoside (Kaempferitrin)	2	m/z cloud
49	12.50	1126.3784	$C_{51}H_{66}O_{28}$	3.87	Kaempferol derivative	3	
60b	16.80	305.0617	$C_{15}H_7D_3O_7$	0.66	Quercetin-d3	1	IS
Flavones							
42	11.30	464.0967	$C_{21}H_{20}O_{12}$	2.55	Tricetin hexoside	2/3	
48b	12.40	842.2140	$C_{36}H_{42}O_{23}$	2.76	Luteolin deoxyhexoside malonylhexoside hexoside	2/3	
50	12.70	448.1017	$C_{21}H_{20}O_{11}$	2.50	Luteolin-7-*O*-glucoside	1	
51	12.90	626.1501	$C_{27}H_{30}O_{17}$	2.79	Tricetin di-hexoside	2/3	
52	13.55	464.0968	$C_{21}H_{20}O_{12}$	2.81	Tricetin hexoside	2/3	
53	14.00	594.1591	$C_{27}H_{30}O_{15}$	1.03	Luteolin deoxyhexoside hexoside	2/3	
54b	14.35	610.1550	$C_{27}H_{30}O_{16}$	2.63	Tricetin deoxyhexoside hexoside	2/3	
55	14.50	448.1013	$C_{21}H_{20}O_{11}$	1.68	Luteolin 4'-*O*-glucoside	1	
56b	14.80	534.1018	$C_{24}H_{22}O_{14}$	1.57	Luteolin malonylhexoside	2/3	
57	14.95	448.1013	$C_{21}H_{20}O_{11}$	1.68	Luteolin hexoside	2/3	
58	15.15	464.0966	$C_{21}H_{20}O_{12}$	2.35	Tricetin hexoside	2/3	
59	15.35	302.0431	$C_{15}H_{10}O_7$	1.46	Tricetin	1	
60	16.50	534.1019	$C_{24}H_{22}O_{14}$	1.63	Luteolin malonylhexoside	2/3	
61	16.80	550.0967	$C_{24}H_{22}O_{15}$	1.53	Tricetin malonylhexoside	2/3	
62	17.20	286.0481	$C_{15}H_{10}O_6$	1.15	Luteolin	1	
Other							
1	2.65	242.0903	$C_{10}H_{14}N_2O_5$	0.18	Thymidine	2	m/z cloud
3	3.35	204.0896	$C_{11}H_{12}N_2O_2$	−1.35	Tryptophan	2	m/z cloud
4	3.50	324.0365	$C_9H_{13}N_2O_9P$	1.87	Uridine monophosphate	2	m/z cloud
10	5.10	444.1998	$C_{21}H_{32}O_{10}$	0.67	epi-Dihydrophaseic acid 4'-*O*-glucoside	2	(Tsopmo & Muir, 2010)
11	5.45	322.0566	$C_{10}H_{15}N_2O_8P$	−0.09	Thymidine monosphosphate	2/3	
12b	5.65	596.1386	$C_{26}H_{28}O_{16}$	1.43	Delphinidin 3-*O*-(2-*O*-β-D-Glucopyranosyl-α-l-arabinopyranoside)	2	(Takeoka et al., 2005)
14	6.00	246.1370	$C_{14}H_{18}N_2O_2$	0.58	Hypaphorine	2	(Tsopmo & Muir, 2010)
32b	9.50	414.1171	$C_{18}H_{22}O_{11}$	2.07	Formononetin rhamnoside	2	(Guo et al., 2022)
34	9.90	320.1009	$C_{15}H_{16}N_2O_6$	0.27	Tryptophan derivative	3	
35b	10.05	402.1895	$C_{19}H_{30}O_9$	1.19	Trihydroxy-megastigmadien-one hexoside	2	(Farag et al., 2019)
36	10.35	197.0771	$C_{10}H_7D_3O_4$	2.01	Ferulic acid – D3	1	IS
37	10.55	492.1852	$C_{21}H_{32}O_{13}$	1.78	Formononetin derivative	3	
43	11.50	390.1323	$C_{20}H_{22}O_8$	2.23	Resveratrol 3-*O*-glucoside (polydatin)	1	
54	14.30	184.0005	$C_7H_4O_6$	−1.57	Chelidonic acid	2	(Khallouki et al., 2018)
56	14.75	234.0989	$^{13}C_6C_8H_{12}O_3$	0.68	Resveratrol (4-hydroxyphenyl-$^{13}C_6$)	1	IS
Unidentified							
26	8.25	530.2008	$C_{24}H_{34}O_{13}$	1.64	Unknown malonyl hexoside	4	
46	12.15	556.1230	$C_{27}H_{24}O_{13}$	2.34	Unknown hexoside	4	

ŋ: This is putatively identified as a mixture of two isomers as reported by (Taylor et al., 2007). The full names of the isomers are: kaempferol 3-*O*-β-glucopyranosyl (1→2)-*O*-[α-rhamnopyranosyl(1→6)-β-galactopyranoside-7-*O*-α-rhamnopyranoside and kaempferol 3-*O*-β-glucopyranosyl(1→2)-*O*-[α-rhamnopyranosyl(1→6)-β-glucopyranoside-7-*O*-α-rhamnopyranoside.

†: One of these two compounds is Kaempferol 3-*O*-β-D-glucopyranosyl(1→2)]-*O*-β-D-galactopyranoside-7-*O*-α-L-rhamnopyranoside, the other is an isomer.

§: One of these two compounds is Kaempferol 3-*O*-{[(6-*O*-*E*-p-coumaroyl)-β-D-glucopyranosyl(1→2)]-α-L-rhamnopyranosyl(1→6)}-β-D-galactopyranoside-7-*O*-α-L-rhamnopyranoside, the other is an isomer.

IS indicates the compound was an internal standard added to the extraction solvent.

not detecting important compounds). Furthermore, many of the polyphenols identified in Table 1, but not included in the targeted method, do not have authentic standards (e.g., prodelphinidins), thereby preventing their absolute quantification. Thus, comparisons in the remainder of this study used relative quantification from the untargeted analysis (as described in Section 2.4.2) as this allows for a comparison of the relative levels of these 85 compounds among these different species.

Since all 85 polyphenols were detected in negative mode (80 in positive mode), this section will focus on the CD 3.3 analysis of the data acquired in negative mode. Using the 85 polyphenols identified in Table 1 (i.e., not including ISs), the PCA results are summarized in Figs. 3A**&B**. The PCA plot in Fig. 3A shows grouping of the individual species, with the notable exception of *L. tomentosus* and *L. orientalis*, which significantly overlap suggesting polyphenol profiles within these two species are the most closely related. In addition, there is a small overlap of this group with *L. culinaris*. All three of these species were identified as being within the primary gene pool using genotyping by sequencing (Wong et al., 2015). Fig. 3B shows a loadings plot for the

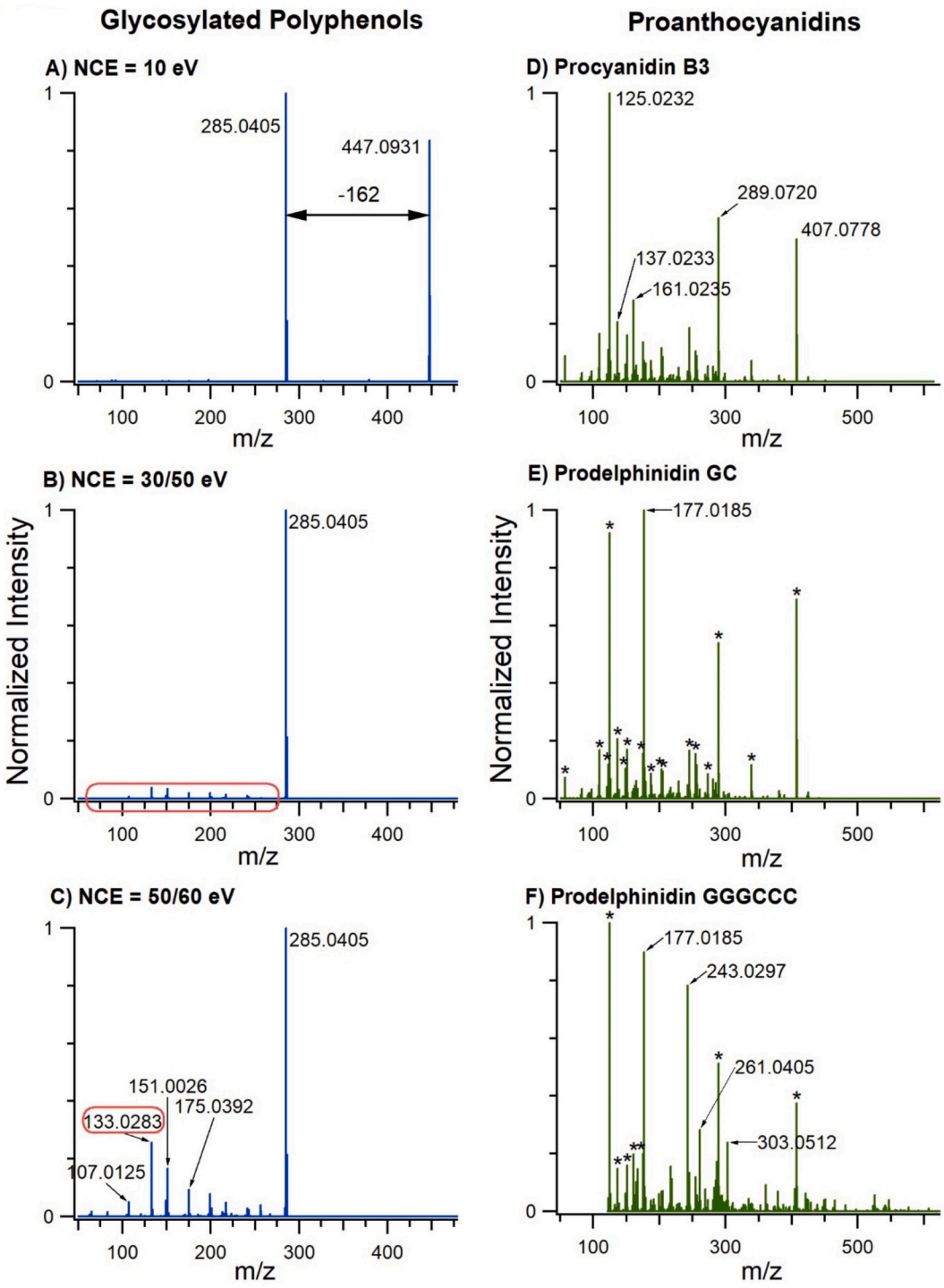

Fig. 2. Examples of a level 2/3 identification. Example 1 shows MS/MS spectra for an unknown glycosylated polyphenol (peak #57 in Table 1) using (A) a normalized collision energy (NCE) of 10/20 eV, (B) NCE = 30/50 eV, and (C) NCE = 50/60 eV. The red circled area shows the fingerprint region in (B) and the red circled m/z value in (C) distinguishes luteolin from kaempferol. Example 2 shows MS/MS spectra using NCE = 30/50 eV for (D) Procyanidin B3 (peak #20 in Table 1), (E) prodelphinidin GC, and (F) prodelphinidin GGGCCC. C refers to (*epi*)catechin monomers and G to (epi)gallocatechin monomers. Peaks in (E) and (F) with the same m/z values as (D) are labelled with *. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

PCA plot in which the numbers correspond to the 85 polyphenols given in Table 1. **Supplementary Fig. S2** shows a PCA plot using the 80 polyphenols detected in positive mode (five hydroxybenzoic acids were not detected). Although the results are mostly similar, some overlap of *L. odemensis* with other species in the primary gene pool is observed, which is believed to be due primarily to the absence of the five hydroxybenzoic acids in this analysis. For example, dihydroxybenzoic acid pentoside hexoside (compound #20b) is not detected in positive mode, but was over an order of magnitude lower in intensity using negative mode in *L. odemensis* compared with the species in the primary gene pool.

As described in Section 2.5.3, a more extensive analysis of the

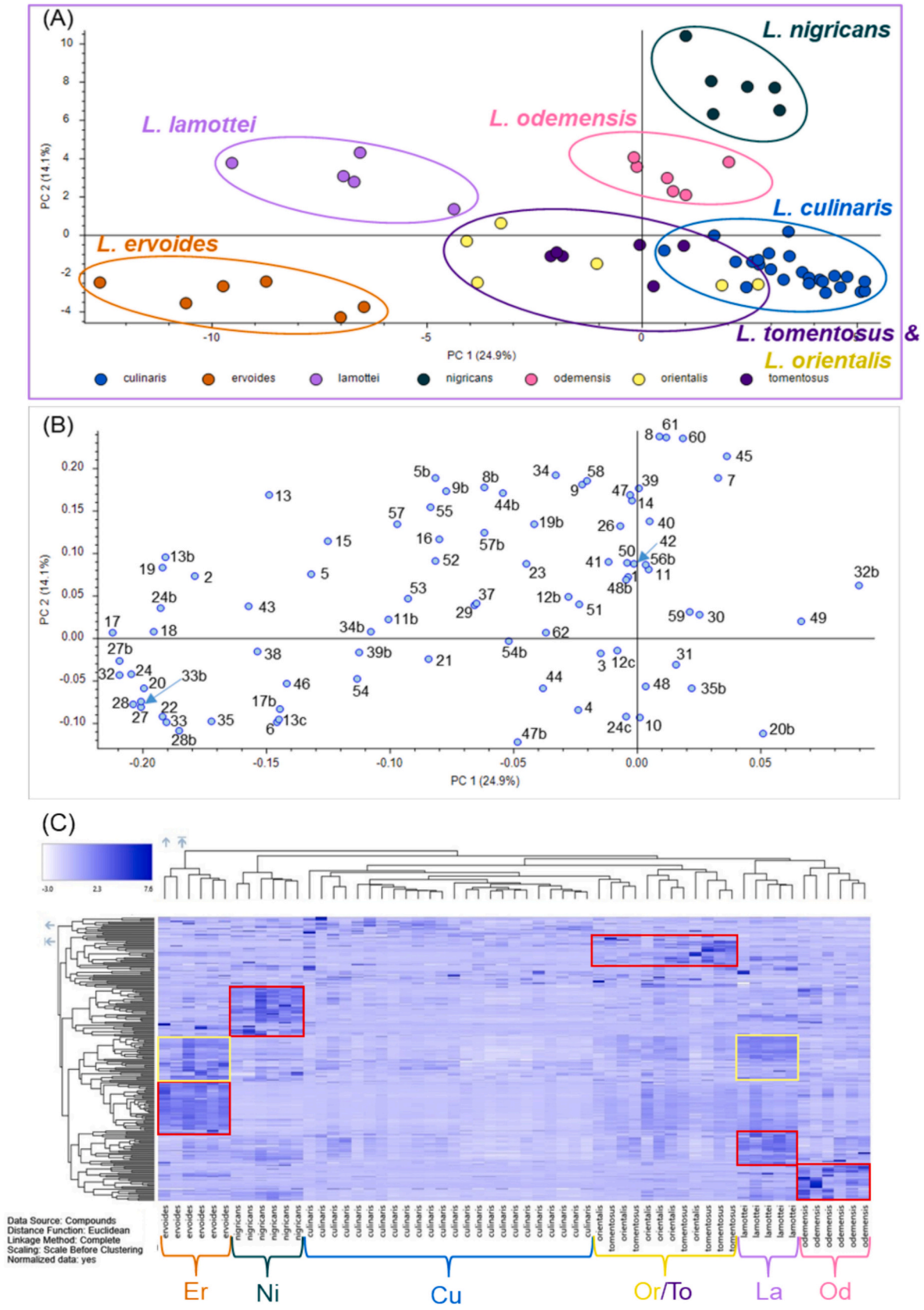

Fig. 3. Statistical analysis of negative mode LC-MS data using Compound Discoverer 3.3. (A) PCA plot (PCA1 vs. PCA2) using the 85 polyphenols in Table 1. (B) Loadings plot with the numbers corresponding to those in Table 1, (C) Hierarchical clustering analysis (HCA) of 231 polyphenols, highlighting clusters of interest within individual species (red square) and a shared cluster of interest (yellow square). Cu = *L. culinaris*, Or = *L. orientalis*, To = *L. tomentosus*, Od = *L. odemensis*, La = *L. lamottei*, Er = *L. ervoides*, and Ni = *L. nigricans*. Note that *L. orientalis* and *L. tomentosus* are overlapping (Or/To).

polyphenols using CD 3.3 identified 231 polyphenols including those given in Table 1. Of these 231 polyphenols, which includes all the entries in Table 1, **Supplementary Table S3** (those having an RT value), and **Supplementary Table S4**, only two were unidentified (level 4) in Table 1; the other 229 were identified at level 3 or higher. These polyphenols, which will be examined in more detail in Section 3.4.3, include 111 proanthocyanidins, 14 flavan-3-ols, 34 phenolic acids, 31 flavonols, 19 flavones, and 20 classified as other. These 231 polyphenols were used to generate the HCA plot shown in Fig. 3C. The x-axis in this plot shows that the genotypes for each species largely group together in the plot, except for *L. tomentosus* and *L. orientalis*, which group together. The squares in the plot are used to highlight significant clusters; the 2 yellow squares represent a shared cluster, whereas the five red squares represent clusters higher in one species. The red cluster in *L. ervoides* consists of 40 polyphenols, of which 32 are proanthocyanidins. Interestingly, these proanthocyanidins consist of mostly C monomers (recall catechin/epicatechin) as this group includes 24 procyanidins and only eight prodelphinidins, with the prodelphinidins containing only one G (recall gallocatechin/epigallocatechin) in the oligomer. The red cluster identified in *L. lamottei* also contains a large number of proanthocyanidins, 20 out of 29 polyphenols. However, these are all prodelphinidins and they mostly contain G monomers. The yellow shared cluster between *L. lamottei* and *L. ervoides* also has all prodelphinidins, but there is approximately an equal mix of G and C monomers. The *L. nigricans* cluster also has a significant amount of proanthocyanidins (23 in the cluster of 42 compounds), again many prodelphinidins containing mostly G monomers, some having only G monomers. There are also five phenolic acids as well as four flavonols and four flavones in the *L. nigricans* cluster. The red cluster in *L. odemensis* consists of a mixture of similar amounts of procyanidins, prodelphinidins, flavonols and flavones, whereas the *L. tomentosus/orientalis* cluster contains primarily phenolic acids, flavonols and flavones. Although there are a couple of compounds at a higher level in some *L. culinaris* genotypes, the plot shows the lack of any significant *L. culinaris* cluster. In fact, many *L. culinaris* genotypes show large regions of light colour, which indicates relatively low polyphenol intensities compared with the wild species. The results in this section show that the PCA and HCA analyses effectively group lentil species based on polyphenol profiles, supporting the genetic relationships between species,

3.4. Differences in compound classes among lentil species

3.4.1. Proanthocyanidins

Proanthocyanidins, formed via the flavonoid biosynthetic pathway, are major contributors to antioxidant capacity in lentil seed coats (Elessawy et al., 2021). They are the most abundant polyphenol class in the lentil seed coat (Dueñas et al., 2003) and this is illustrated in Fig. 4. Fig. 4A shows a comparison of the LC-PDA spectra for the ID samples of *L. ervoides*, *L. culinaris*, and a solvent blank. The method blank illustrates how the absorbance changes with the gradient, which doesn't explain the elevated baseline in the absorbance between ~4 and 16 min in the samples. The elevated baseline in the samples is due to numerous overlapping proanthocyanidins; the UV peaks labelled in Fig. 1 are imposed on top of this broad peak. Among the lentil species, the broad peak is the highest in *L. ervoides* and the lowest in *L. culinaris*. Fig. 4B shows a sum of the full scan mass spectra from 4 to 16 min using the *L. ervoides* ID sample in negative mode. The figure shows numerous proanthocyanidin peaks with distinct, predictable regions of oligomers in the mass spectra. In the figure, the dimers, trimers, and tetramers are singly charged, whereas the pentamers and higher order oligomers are doubly charged. The nonamers ($n = 9$) are clearly visible, and even higher order oligomers (e.g., $n = 11$ shown in the inset) are present at lower abundances (inset of Fig. 4B). Although other polyphenol peaks are present (i.e., from Table 1), the mass spectrum in Fig. 4B is largely dominated by several different procyanidins and prodelphinidins. Within each oligomer region, singly charged peaks are separated by 16 *m/z* units (difference between G and C oligomers), and as will be discussed, multiple oligomers are typically observed for each m/z value.

Thus, proanthocyanidins are a major component of lentils as seven procyanidins and 19 prodelphinidins were identified in Table 1, and in addition to these major species, there are numerous less abundant proanthocyanidins that are present in lentils. **Supplementary Table S4** lists many of these less abundant components and illustrates how they were readily identified. **Supplementary Table S4** shows the retention times of the various isomeric species present for a given oligomer, and the table contains 111 proanthocyanidins and 14 flavan-3-ols. GCC trimers showed the largest number of isomeric species (eight isomers) detected. Fig. 4C shows a limitation of identifying individual isomers as the oligomers increase in length. A plot of GCCCCCC oligomers (prodelphinidin heptamer), shows a distinct peak at 9.8 min, but as there are numerous lower intensity isomers that overlap (the amount increases with oligomer length), selecting individual isomers is no longer feasible. For this reason, **Supplementary Table S4** stops at heptamers, with the exception of five large and distinct octamers. Fig. 4C and the large number of peaks in **Supplementary Table S4** indicates that there are likely hundreds, if not thousands, of proanthocyanidins that make up the baseline shown in Fig. 4A. As the proanthocyanidins are most abundant in wild relatives, crosses of these relatives with cultivated lentils could enhance the bioactive polyphenols content in the latter.

Figs. 4D-4F illustrate how the proanthocyanidin dimers differ among the lentil species. The traces are arranged such that the three species in the primary gene pool are shown at the bottom, the two species in the secondary gene pool in the middle, and the two species in the tertiary and quaternary gene pools at the top. For the CC dimer in Fig. 4D, there is a large peak at ~7.1 min for most lentil species which corresponds to Procyanidin B3 (peak #20 in Fig. 1). This peak is largest in *L. ervoides*, and the smallest in both *L. nigricans* and *L. culinaris*. Although the abundances of the CC dimers are different in the primary and secondary gene pools, the relative ratios of the individual isomers are very similar. There are slight differences in the relative ratios of the isomers for *L. ervoides* (tertiary gene pool) compared with the primary gene pool, but much bigger differences are found in *L. nigricans* (quaternary gene pool). Fig. 4E shows the GC dimer, which is least abundant in the primary gene pool, and most abundant in *L. lamottei*. *L. nigricans* has much more GC than CC when compared with the other species. In Fig. 4F, which shows the less abundant GG dimer, *L. nigricans* has four isomers, each one is the most abundant among the species. Thus, from the HCA plots and the results above, the genetic variation influences the proanthocyanidin profiles, especially for the relative amounts of G/C monomers that are very different in *L. nigricans*. These deviations are consistent with genetic studies that revealed that *L. nigricans* was the most distinct species from the other *Lens* species (Liber et al., 2021; Wong et al., 2015).

3.4.2. Flavan-3-ols

Out of the 231 polyphenols identified with CD 3.3, 14 were flavan-3-ols, 11 of which were glycosylated. Catechin 3-*O*-glucoside is the most abundant flavan-3-ol in all the species. In general, the largest amounts of the flavan-3-ols are found in *L. ervoides,* followed by *L. lamottei*, especially for catechin 3-*O*-glucoside. Conversely, as is shown in Fig. 5A, although the isomer at 4.6 min is the highest for most of the species, there are also other significant isomers of glycosylated (epi)gallocatechin. Fig. 5A also shows that the primary and secondary gene pools show similar traces, except *L. culinaris,* which has a uniquely intense isomer near 7 min. The largest differences in relative abundances are observed within the furthest related gene pools as *L. ervoides* and *L. nigricans* show very different distributions compared with the other lentil species. Consistent with what was observed for the relative amounts of the C and G monomers in the proanthocyanidins, the glycosylated (epi)gallocatechins are most abundant in *L. nigricans*, whereas the glycosylated (epi)catechins are most abundant in *L. ervoides*.

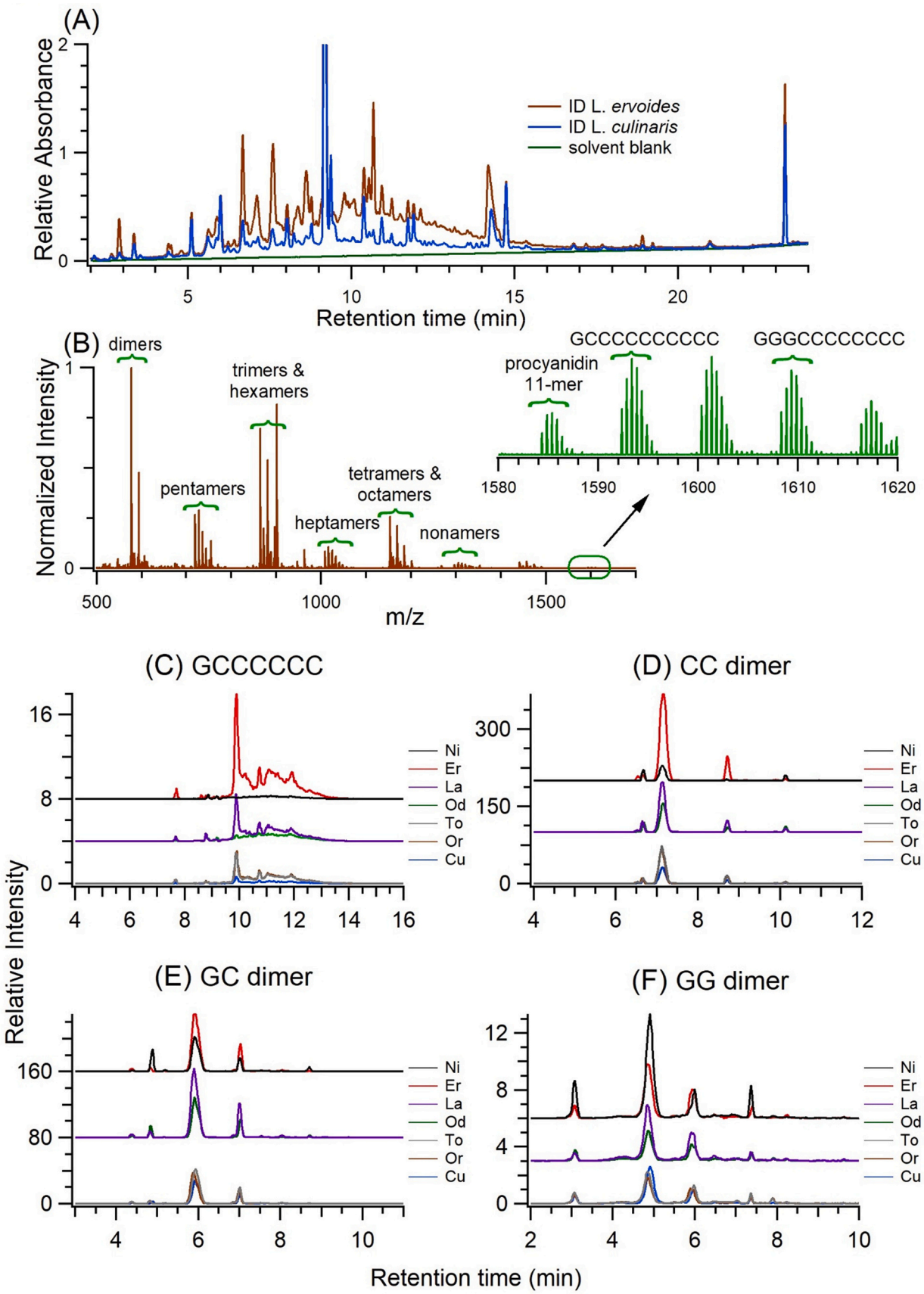

Fig. 4. Proanthocyanidins. (A) Overlaid LC-PDA plots for the ID samples for *L. ervoides* and *L. culinaris,* and a solvent blank. (B) Sum of all mass spectra for the *L. ervoides* ID sample between 4 and 16 min with the inset showing procyanidin and prodelphinidin 11-mers. Plots of m/z vs. retention time to show isomers of (C) prodelphinidin GCCCCCC hepatmers, (D) procyandin CC dimers, (E) GC dimers, and (F) GG dimers. Cu = *L. culinaris*, Or = *L. orientalis*, To = *L. tomentosus,* Od = *L. odemensis*, La = *L. lamottei*, Er = *L. ervoides*, and Ni = *L. nigricans.*

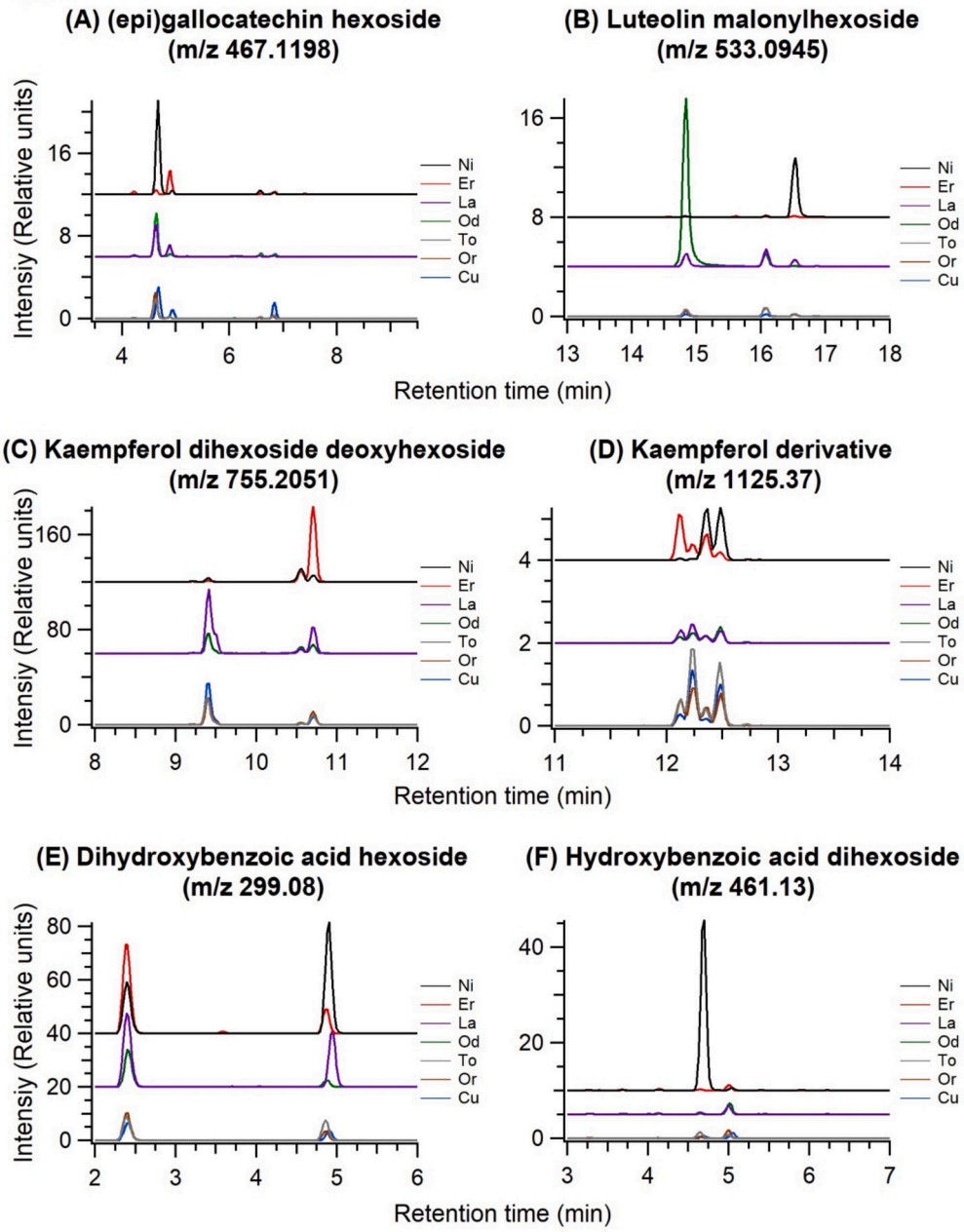

Fig. 5. Comparing isomeric structures among the *Lens* species. For (A) epi(gallocatchin) hexoside, (B) Luteolin malonylhexoside, (C) Kaempferol di-hexoside deoxyhexoside, (D) Kaempferol derivative, (E) Dihydroxybenzoic acid hexoside, (F) Hydroxybenzoic acid di-hexoside. Cu = *L. culinaris*, Or = *L. orientalis*, To = *L. tomentosus,* Od = *L. odemensis*, La = *L. lamottei*, Er = *L. ervoides*, and Ni = *L. nigricans.*

3.4.3. Flavones

The flavones elute toward the end of the LC run, and of the 19 flavones that were included in the 231 polyphenols used to generate Fig. 3B, 17 were glycosylated, and 18 contained either luteolin or tricetin. Some noticeable differences in the profiles among the lentil species are observed, and an example is shown in Fig. 5B. Three main isomers of luteolin malonylhexoside are observed, and the distributions are very different when comparing *L. nigricans*, *L. ervoides*, and *L. odemensis*, whereas the distribution of these isomers for the other four species are similar. In addition to the virtually unique flavones found in *L. nigricans* (peak #60, #61 in Fig. 1), which highlights the diversity of polyphenols available for breeding programs, the more abundant flavones were highest in either *L. lamottei, L. tomemtosus,* and *L. orientalis* (peaks #52, #53, #55 in Fig. 1). In general, the relative amount of the flavones was much lower in *L. ervoides* and *L. culinaris,* and this was also observed for the example given in Fig. 5B.

3.4.4. Flavonols

There were 31 flavonols identified by CD 3.3, 30 were glycosylated, and the majority contained either kaempferol (21 derivatives) or quercetin (8 derivatives). The most abundant flavonols are the kaempferol containing compounds, as can be seen from Table 1. The dominant peak in Fig. 1 (also see **Supplementary Fig. S1**) at 9.2 min (peak 30) is consistent with a kaempferol containing 2 deoxyhexosides and 2 hexosides (and will be referred to as kaempferol di-deoxyhexoside di-hexoside). The peak is putatively assigned in Table 1 since a previous study reported this peak to contain a mixture of kaempferol 3-*O*-β-glucopyranosyl(1→2)-*O*-[α-rhamnopyranosyl(1→6)-β-galactopyranoside-7-*O*-α-rhamnopyranoside and kaempferol 3-*O*-β-glucopyranosyl(1→2)-*O*-[α-rhamnopyranosyl(1→6)-β-glucopyranoside-7-*O*-

α-rhamnopyranoside (Taylor et al., 2007; Tsopmo & Muir, 2010). A previous study found that kaempferol containing species were present at a much higher concentration in the embryo of lentils than the seed coat, whereas all other polyphenols that were investigated were greater in the seed coat (Mirali et al., 2016). Since the embryo represents the majority of the seed and the seed coat is only ~10 % of the seed weight, the dominance of kaempferol di-deoxyhexoside di-hexoside can be partly attributed to its abundance in the embryo.

Kaempferol is generally highly glycosylated and different isomers show varying relative ratios across the species, although again, the largest differences are in the tertiary and quaternary gene pools. This is illustrated in Figs. 5C-5D, which show plots for the isomers of two different kaempferol species. In both these figures, the relative abundances of the isomers in *L. ervoides* and *L. nigricans* are much different compared with the other species.

3.4.5. Phenolic acids

Out of the 34 phenolic acids identified using CD 3.3, the vast majority were hydroxybenzoic acids as only one hydroxycinnamic acid, coumaric acid hexoside, was detected. In addition, 30 of the phenolic acids were glycosylated. Many of the most abundant of these acids were present in similar amounts across the species, although some exceptions were observed. For example, Fig. 5E shows the isomers for dihydroxybenzoic acid hexoside across the species. Careful inspection of the area near 5 min shows two nearly overlapping isomers, thereby making 3 total. This observation highlights a difference for *L. lamottei* and *L. nigricans* compared with the other species. Lower abundance phenolic acids did show more variation, and as was mentioned previously, these acids were detected either as sodium adducts or not at all in positive mode. Thus, differences among the hydroxybenzoic acids, such as for dihydroxybenzoic acid pentoside hexoside (peak #20b) were not properly accounted for in the PCA plot for positive mode, and the additional overlap in that plot (**Supplementary Fig. S2**) highlights the importance of the phenolic acids in examining differences among the species. Note that removing the phenolic acids from the PCA plots in negative mode resulted in similar additional overlap as was observed in **Supplementary Fig. S2**. Fig. 5F also shows an example of a lower abundance hydroxybenzoic acid containing two hexose groups. The plot shows that this compound is much more intense in *L. nigricans* than it is in the other species. Overall, Figs. 4 and 5 show that many polyphenols exist in multiple isomeric forms and the profile and intensity of these isomers does vary among lentil species. Further investigation to better understand these isomers is needed to fully utilize the diversity within the lentil species, since isomers will influence biological activity, chemical reactivity, and interactions of polyphenols.

3.5. Further exploring differences among species

Supplementary Fig. S3 shows comparisons using differential analysis to highlight significant differences in compound regulation; **Supplementary Figs. S3A-F** each show a volcano plot of a wild lentil species compared with *L. culinaris* using the same 231 polyphenols used in generating the HCA plot in Fig. 3C. The plots in the figure go from the least genetically related (*L. nigricans*, **Supplementary Fig. S3A**) to the most closely related (*L. tomentosus,* **Supplementary Fig. S3E**, and *L. orientalis,* **Supplementary Fig. S3F**). All the volcano plots show that the wild lentils have significantly more upregulated compounds (red-highlighted area) versus *L. culinaris*, suggesting higher levels of bioactive compounds in the wild species. The larger number of down-regulated compounds for species in the secondary gene pool and especially the tertiary and quaternary gene pools, is likely due to greater variability in polyphenol distributions, largely caused by isomeric species, as was illustrated in Fig. 5. That is, the primary gene pool is where the isomer profiles of the polyphenol compounds are most similar to *L. culinaris*, whereas more distant gene pools can have very different distributions. For example, one of the down-regulated compounds for *L. nigricans* and *L. ervoides* is kaempferol di-hexoside deoxyhexoside peak at 9.5 min in Fig. 5C, however *L. ervoides* shows a large isomeric peak in the same figure at 10.7 min. When comparing the most closely related species in the primary gene pool, *L. orientalis* and *L. tomentosus*, only one compound in *L. orientalis* is significantly down-regulated, whereas 38 (*L. tomentosus*) and 30 (*L. orientalis*) are significantly up-regulated.

When looking at the significantly upregulated compounds in *L. orientalis* (vs. *L. culinaris*), there are 15 procyanidins, six prodelphinidins (five with only one G), six flavones, two hydroxybenzoic acids, and a quercetin derivative (**Supplementary Fig. S3F**). This result, which is consistent with previous figures, demonstrates that there are relatively low amounts of procyanidins (and flavones) in *L. culinaris* compared with other wild lentils species. **Supplementary Fig. S3G** shows a volcano plot comparing two species in the primary gene pool, *L. orientalis* and *L. tomentosus*. As was shown in the PCA plot (Fig. 3), *L. orientalis* and *L. tomentosus* overlapped, and the volcano plot comparing the 2 species also shows minimal differences as only one compound has changed significantly. These results again suggest that these two species have similar seed polyphenol properties. **Supplementary Fig. S3H** shows a volcano plot comparing the two species in the secondary gene pool, and since only 10 polyphenols are significantly different, this again indicates similar seed polyphenol properties.

4. Conclusions and implications

An investigation of polyphenol profiles among lentil species identified the most abundant polyphenols and revealed significant differences in polyphenol profiles among the lentil species. PCA and HCA analyses effectively grouped the lentil species based on polyphenol profiles, which supports the genetic relationships between species, while the volcano plots identified significant differences among the species. Proanthocyanidins are the major class of polyphenols in lentils with prodelphinidins being more prevalent in *L. nigricans* and procyanidins being more prevalent in *L. ervoides*. The LC-PDA detection of the proanthocyanidins illustrated the lowest amount of these compounds were found in cultivated lentils (*L. culinaris*). Other polyphenol subclasses (e.g., flavan-3-ols, flavones, flavonols, and phenolic acids) also typically showed lower levels in *L. culinaris.* These results suggest that domestication has increased the agronomical performance of lentil but narrowed down its seed quality properties.

In addition, glycosylation was prevalent among flavan-3-ols, flavones, flavonols, and phenolic acids, however, it was rare within the proanthocyanins. The large number of isomeric compounds among the wild relatives shows that the lentil wild relatives harbour natural variation for improved polyphenol concentration. This could provide opportunities to breeders to breed for enhanced seed quality traits in this valuable food protein crop globally and ultimately improve heath benefits of lentil seeds.

Funding

This research was funded by the Agricultural Development Fund, Government of Saskatchewan (ADF, external ID: 20150285); The Natural Sciences and Engineering Research Council of Canada Industrial Research Chair Program (NSERC, external ID: IRCPJ 395994–14/IRCSA 395993–14); Saskatchewan Pulse Growers (SPG, external ID: IRC 386279–09), and Western Grains Research Foundation (external ID: VarD1609). Fatma Elessawy is partially funded by a Natural Sciences and Engineering Research Council of Canada Discovery Grant (external ID: RGPIN-2021-03293).

CRediT authorship contribution statement

Randy W. Purves: Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation,

Conceptualization. **Hamid Khazaei:** Writing – review & editing, Writing – original draft, Resources, Methodology. **Fatma M. Elessawy:** Writing – review & editing, Methodology, Investigation, Formal analysis. **Roger Munro:** Writing – review & editing, Methodology. **Bryn O. Shurmer:** Writing – review & editing, Resources. **Albert Vandenberg:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Thermo Fisher for the use of the TSQ Vantage triple quadrupole mass spectrometer as part of a collaboration between the University of Saskatchewan and Thermo Fisher. The authors also thank Devini De Silva and the Pulse Crop Field Crew for technical assistance with production and seed sourcing of wild lentil samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2025.116154.

Data availability

Data will be made available on request.

References

Aguilera, Y., Dueñas, M., Estrella, I., Hernández, T., Benitez, V., Esteban, R. M., & Martín-Cabrejas, M. A. A. (2010). Evaluation of phenolic profile and antioxidant properties of Pardina lentil as affected by industrial dehydration. *Journal of Agricultural and Food Chemistry, 58*(18), 10101–10108. https://doi.org/10.1021/jf102222t

Barilli, E., & Rubiales, D. (2023). Identification and characterization of resistance to rust in lentil and its wild relatives. *Plants, 12*(3), 626. https://doi.org/10.3390/plants12030626

Bhadauria, V., Wong, M. M., Bett, K. E., & Banniza, S. (2016). Wild help for enhancing genetic resistance in lentil against fungal diseases. *Current Issues in Molecular Biology, 19*(1), 3–6. https://doi.org/10.21775/9781910190357.02

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S. R., & Varshney, R. K. (2022). Reap the crop wild relatives for breeding future crops. *Trends in Biotechnology, 40*(4), 412–431. https://doi.org/10.1016/j.tibtech.2021.08.009

Coyne, C. J., Kumar, S., von Wettberg, E. J., Marques, E., Berger, J. D., Redden, R. J., & Smýkal, P. (2020). Potential and limits of exploitation of crop wild relatives for pea, lentil, and chickpea improvement. *Legume Science, 2*(2), Article e36. https://doi.org/10.1002/leg3.36

Dixon, R. A. (2001). Natural products and plant disease resistance. *Nature, 411*(6839), 843–847. https://doi.org/10.1038/35081178

Dueñas, M., Hernández, T., & Estrella, I. (2002). Phenolic composition of the cotyledon and the seed coat of lentils (*Lens culinaris* L.). *European Food Research and Technology, 215*, 478–483. https://doi.org/10.1007/s00217-002-0603-1

Dueñas, M., Sun, B., Hernández, T., Estrella, I., & Spranger, M. I. (2003). Proanthocyanidin composition in the seed coat of lentils (*Lens culinaris* L.). *Journal of Agricultural and Food Chemistry, 51*(27), 7999–8004. https://doi.org/10.1021/jf0303215

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., & Haselden, J. N. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols, 6*(7), 1060–1083. https://doi.org/10.1038/nprot.2011.335

Elessawy, F. M., Bazghaleh, N., Vandenberg, A., & Purves, R. W. (2020). Polyphenol profile comparisons of seed coats of five pulse crops using a semi-quantitative liquid chromatography-mass spectrometric method. *Phytochemical Analysis, 31*(4), 458–471. https://doi.org/10.1002/pca.2909

Elessawy, F. M., Hughes, J., Khazaei, H., Vandenberg, A., El-Aneed, A., & Purves, R. W. (2023). A comparative metabolomics investigation of flavonoid variation in faba bean flowers. *Metabolomics, 19*(6), 52. https://doi.org/10.1007/s11306-023-02014-w

Elessawy, F. M., Vandenberg, A., El-Aneed, A., & Purves, R. W. (2021). An untargeted metabolomics approach for correlating pulse crop seed coat polyphenol profiles with antioxidant capacity and iron chelation ability. *Molecules, 26*(13), 3833. https://doi.org/10.3390/molecules26133833

Elessawy, F. M., Wright, D., Vandenberg, A., El-Aneed, A., & Purves, R. W. (2023). Mass spectrometry-based untargeted metabolomics reveals the importance of glycosylated flavones in patterned lentil seed coats. *Journal of Agricultural and Food Chemistry, 71*(7), 3541–3549. https://doi.org/10.1021/acs.jafc.2c07844

FAOSTAT. (2024). Food and Agriculture Organization of the United Nations. Retrieved from Available online http://faostat.fao.org accessed on 3 February 2023.

Farag, M. A., Khattab, A. R., Maamoun, A. A., Kropf, M., & Heiss, A. G. (2019). UPLC-MS metabolome based classification of *Lupinus* and *Lens* seeds: A prospect for phyto-equivalency of its different accessions. *Food Research International, 115*, 379–392. https://doi.org/10.1016/j.foodres.2018.11.003

Farkas, G., & Kiraaly, Z. (1962). Role of phenolic compounds in the physiology of plant diseases and disease resistance. *Journal of Phytopathology, 44*(2), 105–150. https://doi.org/10.1111/j.1439-0434.1962.tb02005.x

Grela, E. R., Kiczorowska, B., Samolińska, W., Matras, J., Kiczorowski, P., Rybiński, W., & Hanczakowska, E. (2017). Chemical composition of leguminous seeds: Part I—Content of basic nutrients, amino acids, phytochemical compounds, and antioxidant activity. *European Food Research and Technology, 243*, 1385–1395. https://doi.org/10.1007/s00217-017-2849-7

Gulewicz, P., Ciesiołka, D., Frias, J., Vidal-Valverde, C., Frejnagel, S., Trojanowska, K., & Gulewicz, K. (2000). Simple method of isolation and purification of α-galactosides from legumes. *Journal of Agricultural and Food Chemistry, 48*(8), 3120–3123. https://doi.org/10.1021/jf000210v

Guo, F., Peng, L., Xiong, H., Wang, J., Tsao, R., Peng, X., & Sun, Y. (2022). Free and bound phenolics of Laird lentil (*Lens culinaris*) hulls and the anti-inflammatory activity of their digestive products via crosstalk between NF-κB and Keap1-Nrf2 signaling pathways in HT-29 cells. *Journal of Agricultural and Food Chemistry, 70*(41), 13251–13263. https://doi.org/10.1021/acs.jafc.2c04471

Khallouki, F., Ricarte, I., Breuer, A., & Owen, R. W. (2018). Characterization of phenolic compounds in mature Moroccan Medjool date palm fruits (*Phoenix dactylifera*) by HPLC-DAD-ESI-MS. *Journal of Food Composition and Analysis, 70*, 63–71. https://doi.org/10.1016/j.jfca.2018.03.005

Khazaei, H., Caron, C. T., Fedoruk, M., Diapari, M., Vandenberg, A., Coyne, C. J., & Bett, K. E. (2016). Genetic diversity of cultivated lentil (*Lens culinaris* Medik.) and its relation to the world's agro-ecological zones. *Frontiers in Plant Science, 7*, 1093. https://doi.org/10.3389/fpls.2016.01093

Khazaei, H., Subedi, M., Nickerson, M., Martínez-Villaluenga, C., Frias, J., & Vandenberg, A. (2019). Seed protein of lentils: Current status, progress, and food applications. *Foods, 8*(9), 391. https://doi.org/10.3390/foods8090391

Kumar, S., Choudhary, A. K., Rana, K. S., Sarker, A., & Singh, M. (2018). Bio-fortification potential of global wild annual lentil core collection. *PLoS One, 13*(1), Article e0191122. https://doi.org/10.1371/journal.pone.0191122

Liber, M., Duarte, I., Maia, A. T., & Oliveira, H. R. (2021). The history of lentil (*Lens culinaris* subsp. *culinaris*) domestication and spread as revealed by genotyping-by-sequencing of wild and landrace accessions. *Frontiers in Plant Science, 12*, Article 628439. https://doi.org/10.3389/fpls.2021.628439

Llorach, R., Favari, C., Alonso, D., Garcia-Aloy, M., Andres-Lacueva, C., & Urpi-Sarda, M. (2019). Comparative metabolite fingerprinting of legumes using LC-MS-based untargeted metabolomics. *Food Research International, 126*, Article 108666. https://doi.org/10.1016/j.foodres.2019.108666

Mirali, M., Ambrose, S. J., Wood, S. A., Vandenberg, A., & Purves, R. W. (2014). Development of a fast extraction method and optimization of liquid chromatography–mass spectrometry for the analysis of phenolic compounds in lentil seed coats. *Journal of Chromatography B, 969*, 149–161. https://doi.org/10.1016/j.jchromb.2014.08.007

Mirali, M., Purves, R. W., Stonehouse, R., Song, R., Bett, K., & Vandenberg, A. (2016). Genetics and biochemistry of zero-tannin lentils. *PLoS One, 11*(10), Article e0164624. https://doi.org/10.1371/journal.pone.0164624

Mirali, M., Purves, R. W., & Vandenberg, A. (2017). Profiling the phenolic compounds of the four major seed coat types and their relation to color genes in lentil. *Journal of Natural Products, 80*(5), 1310–1317. https://doi.org/10.1021/acs.jnatprod.6b00872

Mustafa, A. M., Abouelenein, D., Acquaticci, L., Alessandroni, L., Angeloni, S., Borsetta, G., & Vittori, S. (2022). Polyphenols, saponins and phytosterols in lentils and their health benefits: An overview. *Pharmaceuticals, 15*(10), 1225. https://doi.org/10.3390/ph15101225

Nicholson, R. L., & Hammerschmidt, R. (1992). Phenolic compounds and their role in disease resistance. *Annual Review of Phytopathology, 30*(1), 369–389. https://doi.org/10.1146/annurev.py.30.090192.002101

Pandey, K. B., & Rizvi, S. I. (2009). Plant polyphenols as dietary antioxidants in human health and disease. *Oxidative Medicine and Cellular Longevity, 2*(5), 270–278. https://doi.org/10.4161/oxim.2.5.9498

Podder, R., Banniza, S., & Vandenberg, A. (2013). Screening of wild and cultivated lentil germplasm for resistance to stemphylium blight. *Plant Genetic Resources, 11*(1), 26–35. https://doi.org/10.1017/S1479262112000329

Quemener, B., & Brillouet, J.-M. (1983). Ciceritol, a pinitol digalactoside form seeds of chickpea, lentil and white lupin. *Phytochemistry, 22*(8), 1745–1751. https://doi.org/10.1016/S0031-9422(00)80263-0

Rathahao-Paris, E., Alves, S., Junot, C., & Tabet, J.-C. (2016). High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics, 12*, 1–15. https://doi.org/10.1007/s11306-015-0882-8

Singh, M., Kumar, S., Basandrai, A. K., Basandrai, D., Malhotra, N., Saxena, D. R., & Singh, K. (2020). Evaluation and identification of wild lentil accessions for enhancing genetic gains of cultivated varieties. *PLoS One, 15*(3), Article e0229554. https://doi.org/10.1371/journal.pone.0229554

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., & Griffin, J. L. (2007). Proposed minimum reporting standards for chemical analysis:

Chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics, 3*, 211–221. https://doi.org/10.1007/s11306-007-0082-2

Takeoka, G. R., Dao, L. T., Tamura, H., & Harden, L. A. (2005). Delphinidin 3-O-(2-O-β-D-glucopyranosyl-α-L-arabinopyranoside): A novel anthocyanin identified in beluga black lentils. *Journal of Agricultural and Food Chemistry, 53*(12), 4932–4937. https://doi.org/10.1021/jf040493h

Taylor, W. G., Fields, P. G., & Sutherland, D. H. (2007). Fractionation of lentil seeds (*Lens culinaris* Medik.) for insecticidal and flavonol tetraglycoside components. *Journal of Agricultural and Food Chemistry, 55*(14), 5491–5498. https://doi.org/10.1021/jf0705062

Tsopmo, A., & Muir, A. D. (2010). Chemical profiling of lentil (*Lens culinaris* Medik.) cultivars and isolation of compounds. *Journal of Agricultural and Food Chemistry, 58*(15), 8715–8721. https://doi.org/10.1021/jf101412y

Wong, M. M., Gujaria-Verma, N., Ramsay, L., Yuan, H. Y., Caron, C., Diapari, M., & Bett, K. E. (2015). Classification and characterization of species within the genus *Lens* using genotyping-by-sequencing (GBS). *PLoS One, 10*(3), Article e0122025. https://doi.org/10.1371/journal.pone.0122025

Zanotto, S., Bertrand, A., Purves, R. W., Olsen, J. E., Elessawy, F. M., & Ergon, Å. (2023). Biochemical changes after cold acclimation in Nordic red clover (Trifolium pratense L.) accessions with contrasting levels of freezing tolerance. *Physiologia Plantarum, 175*(4), Article e13953. https://doi.org/10.1111/ppl.13953

Zanotto, S., Khazaei, H., Elessawy, F. M., Vandenberg, A., & Purves, R. W. (2020). Do faba bean genotypes carrying different zero-tannin genes (*zt1* and *zt2*) differ in phenolic profiles? *Journal of Agricultural and Food Chemistry, 68*(28), 7530–7540. https://doi.org/10.1021/acs.jafc.9b07866

Zhang, H., Jha, A. B., De Silva, D., Purves, R. W., Warkentin, T. D., & Vandenberg, A. (2019). Improved folate monoglutamate extraction and application to folate quantification from wild lentil seeds by ultra-performance liquid chromatography-selective reaction monitoring mass spectrometry. *Journal of Chromatography B, 1121*, 39–47. https://doi.org/10.1016/j.jchromb.2019.05.007

Zou, Y., Chang, S. K., Gu, Y., & Qian, S. Y. (2011). Antioxidant activity and phenolic compositions of lentil (*Lens culinaris* var. Morton) extract and its fractions. *Journal of Agricultural and Food Chemistry, 59*(6), 2268–2276. https://doi.org/10.1021/jf104640k

Żuchowski, J., Pecio, Ł., & Stochmal, A. (2014). Novel flavonol glycosides from the aerial parts of lentil (*Lens culinaris*). *Molecules, 19*(11), 18152–18178. https://doi.org/10.3390/molecules191118152