Untargeted metabotyping to study phenylpropanoid diversity in crop plants

Adriana Garibay-Hernández1, Nikolas Kessler2, Anna Maria Józefowicz1, Gözde Merve Türksoy1, Ulrike Lohwasser1, Hans-Peter Mock1

1Leibniz Institute for Plant Genetics and Crop Plant Research, Gatersleben, Germany
2Bruker Daltonik GmbH, Bremen, Germany

Correspondence
Hans-Peter Mock, Leibniz Institute for Plant Genetics and Crop Plant Research, Corrensstraße 3, D-06466 Gatersleben, Germany.
Email: mock@ipk-gatersleben.de

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: FKZ 031A352B

Edited by F. Pescheck

Abstract
Plant genebanks constitute a key resource for breeding to ensure crop yield under changing environmental conditions. Because of their roles in a range of stress responses, phenylpropanoids are promising targets. Phenylpropanoids comprise a wide array of metabolites; however, studies regarding their diversity and the underlying genes are still limited for cereals. The assessment of barley diversity via genotyping-by-sequencing is in rapid progress. Exploring these resources by integrating genetic association studies to in-depth metabolomic profiling provides a valuable opportunity to study barley phenylpropanoid metabolism; but poses a challenge by demanding large-scale approaches. Here, we report an LC-PDA-MS workflow for barley high-throughput metabotyping. Without prior construction of a species-specific library, this method produced phenylpropanoid-enriched metabotypes with which the abundance of putative metabolic features was assessed across hundreds of samples in a single-processed data matrix. The robustness of the analytical performance was tested using a standard mix and extracts from two selected cultivars: Scarlett and Barke. The large-scale analysis of barley extracts showed (1) that barley flag leaf profiles were dominated by glycosylation derivatives of isovitexin, iso-orientin, and isoscoparin; (2) proved the workflow’s capability to discriminate within genotypes; (3) highlighted the role of glycosylation in barley phenylpropanoid diversity. Using the barley S42IL mapping population, the workflow proved useful for metabolic quantitative trait loci purposes. The protocol can be readily applied not only to explore the barley phenylpropanoid diversity represented in genebanks but also to study species whose profiles differ from those of cereals: the crop Helianthus annuus (sunflower) and the model plant Arabidopsis thaliana.

1 | INTRODUCTION

The genetic diversity represented in plant germplasm collections constitutes a key resource for breeding plants to ensure crop yield and nutritional quality under changing environmental conditions (Esquinas-Alcazar, 2005; Fowler & Hodgkin, 2004; Gollin, 2020; Vreugdenhil et al., 2005). Several cereals are major staple crop species worldwide, amongst them wheat, rice, maize, and barley. Hence, conservation and evaluation of cereal germplasm collections are of particular relevance for future human and animal nutrition. The significance of the preservation efforts has led to the establishment of plant genebanks in many countries and, more...
recently, the Svalbard global seed vault (Asdal & Guarino, 2018; Westengen et al., 2020). Whereas the maintenance of the large germplasm collections constitutes a continuous effort itself, evaluation of the stocks is still a major bottleneck toward their use for breeding and biotechnology since these demand large-scale approaches. With the development of novel sequencing technologies and the so-called “omics” strategies for high-throughput analysis of transcripts, proteins, and metabolites, the evaluation of the diversity within genebank collections has gained momentum but remains challenging. Large-scale phenotyping of collections is now enabled by the development of systems for automated assessment of plants (Yang et al., 2020). Genomic sequences have been published for the major crop species rice, barley, maize, and wheat (Mascher et al., 2017; Schnable, 2015; Song et al., 2018; Walkowiak et al., 2020). Being amongst these agriculturally important crops, the assessment of the genetic diversity within barley by genomic sequencing is now in rapid progress. Almost all barley accessions of the German ex-situ genebank were analyzed using a genome-wide genotyping-by-sequencing approach; this enabled the detection of known and novel loci underlying morphological traits that differentiate the barley gene pools (Milner et al., 2019). Twenty barley genotypes, including landraces, cultivars, and a wild relative, were used to establish the barley pan-genome representing its global diversity (Jayakodi et al., 2020). These sequencing approaches have resulted in databases such as BRIDGE (König et al., 2020) and barleyVarDB (Tan et al., 2020) that enable a comprehensive analysis of the genomic variation across barley accessions. Altogether, the increasing availability of these resources is consolidating barley as an attractive model not only for crop research but for additional plant science fields (Harwood, 2019), such as plant development (Poursarebani et al., 2020; Walla et al., 2020), chloroplast function (Li et al., 2019; Rotaserti et al., 2020), and plant pathology (Dömmann et al., 2014; Lenk et al., 2018).

The plant specialized metabolism comprises around 200 000 compounds synthesized through a wide array of metabolic pathways from numerous plant taxa (Arimura & Maffei, 2016). Phenylpropanoids contribute to the specialized metabolism with an enormous set of metabolites displaying intermediates of the shikimate pathway as core structures. These are amplified and subjected to additional chemical modifications (e.g., glycosylation, methylation, acylation), resulting in structurally and functionally diverse phenolics that are defined in a species-, tissue-, and temporal-specific manner (Vogt, 2010). Unlike dicotyledonous species such as the model Arabidopsis thaliana (Arabidopsis), whose phenylpropanoids are dominated by flavonol derivatives, cereals namely comprise flavonines (Tohge et al., 2017). Early studies profiling nearly 1500 barley varieties by using TLC (thin-layer chromatography) already highlighted the remarkable variability of flavonoids dominated by C-glucoflavone derivatives (Fröst et al., 1975). The recent annotation of a large number of phenylpropanoids identified in nine barley cultivars using in-depth analytical approaches supports these early observations (Piasecka et al., 2015). Amongst flavonoids, saponarin, the 7-O-glucoside of the 6-C-glucoflavone isovitexin, has been largely reported as the major feature in barley leaves (Brach et al., 2018; Kaspar et al., 2010; Seikel & Geissman, 1957). In contrast, proanthocyanidins (condensed tannins) and their catechin monomers have been shown to accumulate in the seed coats of mature grains (Aastrup et al., 1984). Acylated polyamines such as hydroxyxycinnamoylagmatines and their condensation products, hordatines, as well as hydroxycinnamic acid derivatives (glucosides and quinic acid esters), are also detected in barley (Piasecka et al., 2015; Pihlava, 2014; Stoessl & Unwin, 1970).

Specific phenylpropanoid patterns impact the plant interactions with the environment. However, the knowledge underlying their diversity, specific regulation, and physiological roles remain limited in cereals. The induction of saponarin, lutonarin (isorientin 7-O-glucoside), and feruloylquinic acid in young barley leaves exposed to high photosynthetically active radiation, especially to UV, support their role in photoprotection against high radiation stress (Kaspar et al., 2010; Klem et al., 2015; Reuber et al., 1996). Ferulic and sinapic acid derivatives, acylated flavone glycosides, and hordatines, were found as drought-responsive in leaves from young barley seedlings (Piasecka et al., 2017). Phenylpropanoid accumulation upon increasing CO2 concentrations has also been observed (Hunt et al., 2021). Induction of specific phenylpropanoid branches during cold acclimation has been reported for winter cereals (NDong et al., 2003) and other plant species (Petridis et al., 2016; Solecaka & Kacperska, 2003). Roles in resistance to the barley leaf epidermal fungal pathogen Erysiphe graminis f. sp. hordel have been attributed to p-coumaroyl-hydroxyagmatine, but not to its non-hydroxylated derivative p-coumaroyl-agnatin (von Röpenack et al., 1998). Inhibition of fungal spore germination has been shown for hordatines (Stoessl & Unwin, 1970). Barley mutants accumulating dihydroquercetin in the testa layer of developing caryopses have proven resistance to Fusarium infection (Skadhauge et al., 1997). Also, the significance of particular phenylpropanoids for human nutrition is discussed (Hoensch & Oertel, 2015; Neelam et al., 2020). The analysis of the so-called proanthocyanidin-free Ant mutants isolated since 1977, together with additional studies, have enabled the identification of candidate phenylpropanoid metabolism genes in barley (Peukert et al., 2013; von Wettstein, 2007); phenylalanine ammonia lyase (Kervinen et al., 1997), cinnamate 4-hydroxylase, chalcone synthase (Christensen, Gregersen, Schroder, & Collinge, 1998b), chalcone isomerase (Druka et al., 2003), flavanone 3-hydroxylase (Meldgaard, 1992), flavonoid 3’- and 3,5’-hydroxylases (Vikhorev et al., 2019), dihydroflavonol reductase (Kristiansen & Rohde, 1991), leucoanthocyanidin reductase, leucoanthocyanidin dioxygenase, flavonoid 7-O-methyltransferase (Christensen et al., 1998a), and a glucosyltransferase (Brach et al., 2018). Only a few of these genes have been characterized in biochemical and/or physiological terms. Despite these findings, the knowledge on barley phenylpropanoid metabolism is yet fragmentary and comprehensive studies are necessary.

The analysis of the plant’s specialized metabolism, including phenylpropanoids, still faces major technical obstacles directly associated with their structural complexity, their wide dynamic range, which is highly dependent on environmental conditions, and their complex spatiotemporal dynamics (Li & Gaquerel, 2021). The flexibility of LC–MS (Liquid Chromatography coupled to MS) has resulted in its widespread application for metabolomics studies, being one of the preferred analytical tools for specialized metabolites. However, some
obstacles remain. Additional to their high complexity, their limited coverage in databases not only makes their analysis but also their annotation difficult (Döll et al., 2019). Owing to the multidimensional complexity of the generated data, further processing remains challenging and computationally intensive (Döll et al., 2019). Ongoing developments on MS metabolomics and computational tools are addressing these obstacles and revolutionizing the field (Li & Gaquerel, 2021).

The increasing availability of genomic sequence information for many accessions of a crop species, including barley, and the progress in phytochemical analysis promote integrating genetic mapping and metabolic network analysis. The combination of both approaches will allow identifying novel mechanisms of regulatory control of plant metabolism. Furthermore, the information will allow us to find yet unknown genes involved in the biosynthesis of specialized metabolites in crops. The integration of these strategies is already showing progress in the field. A metabolic genome-wide association study with Arabidopsis accessions grown under two environmental conditions allowed the detection of 123 mQTLs (metabolic Quantitative Trait Loci) and the identification of 70 candidate genes involved in the specialized metabolism (Wu et al., 2018). A similar combinatorial approach was applied to assess Tibetan hulless barley (quingke) and elite barley accessions (Zeng et al., 2020). When barley is cultivated on the Tibetan Plateau, it is exposed to strong UV-B radiation. The study revealed different branches of phenylpropanoid metabolism as relevant for the adaptation toward UV-B (Zeng et al., 2020). In a previous study, we applied LC coupled to UV/Vis-based detection to profile the barley introgression lines (IL) of the S42IL population, which is derived from the introgressive allele hybridization of the Israeli wild barley accession ISR42-8 into the gene pool of the German spring cultivar Scarlett. This study identified a candidate gene encoding a glycosyltransferase involved in the biosynthesis of the flavonoid isovitexin 2’-O-6-D-glucoside (Brauch et al., 2018). We now aimed to extend our earlier approach (Brauch et al., 2018) to enable the comparative phenylpropanoid profiling of larger populations and sets of genotypes potentially comprising several hundreds of samples.

In this study, we report the development of a workflow for the high-throughput untargeted metabolotyping of phenylpropanoids applicable for barley and other plant species. Without the prior construction of a species-specific compound library, this workflow aimed at the generation of metabolic signatures enriched in phenylpropanoids, with which the abundance of thousands of putative metabolic features could be assessed across hundreds of samples in a single-processed data matrix. By pursuing this goal, the challenge of generating and processing large data matrices was undertaken. The workflow was designed to reflect the samples’ metabolic complexity and enable their discrimination in a genotype-, development-, or treatment-specific manner through the analysis of both shared and specific metabolic features. The analytical performance was initially assessed with commercial standards and extracts from two German spring barley elite cultivars, Barke, and Scarlett. The feasibility of using the relative abundances of the identified features as quantitative traits for genetic association studies was evaluated using the S42IL population. Finally, we tested the workflow’s versatility through the analysis of extracts from dicotyledonous species whose phenylpropanoid profiles differ from those of cereals: the crop Helianthus annuus (sunflower) and the model plant Arabidopsis that has largely contributed to current knowledge on specialized metabolism (D’Auria & Gershenzon, 2005; Tohge et al., 2017).

2 | MATERIALS AND METHODS

2.1 | Reagents

Chemicals were purchased from Merck KGaA, Carl-Roth GmbH & Co. KG, and Th. Geyer GmbH & Co. KG. Solvents (acetonitrile, methanol, water) were LC–MS grade (≥ 99.95%; Chempolite®, Th. Geyer). Analytical standards are described in Table S1.

2.2 | Standard mix preparation

A set of eight metabolites (Table 1) comprising soluble single and polyphenolic compounds of different polarities and masses was defined. Starting solutions were prepared at 2 mg mL⁻¹ in LC–MS water (compounds 1–3) or methanol (compounds 4–8). Using these, an initial stock was prepared in 0.5% (v/v) formic acid (final concentration): all compounds were at 50 μg mL⁻¹, except QUE and NEN (5 μg mL⁻¹). To evaluate the workflow performance with the standard compounds, the stock was diluted 1:4 in 80% (v/v) methanol to a final concentration of 10 μg mL⁻¹ for each compound (except QUE and NEN: 1 μg mL⁻¹). Hence, the LC injection of 3 μL corresponded to the analysis of 30 ng of each compound (except QUE and NEN: 3 ng) in-column.

To test the feasibility of using these compounds as an internal standard, barley methanolic extracts were spiked with the stock solution as follows: 20 μL of stock solution added to 80 μL of barley sample extract.

2.3 | Plant materials

2.3.1 | Barley

Seven European spring barley elite cultivars (Hordeum vulgare spp. vulgare cv. Barke, Grace, Propino, Quench, RGT Planet, Salome, Scarlett) and 41 lines from the S42IL population were grown in 2016 under field conditions at the experimental station ‘Merbitz’, Germany (51° 36’ 38.6” N, 11° 53’ 27.8” E; sandy loam soil; annual temperature: 8.8°C; annual precipitation: 450 mm) (Zahn et al., 2020). The S42IL population comprises 83 lines covering 94.5% of the genome from the wild barley ISR42-8 (H. vulgare spp. spontaneum) (Honsdorf et al., 2017). The subset of 41 S42ILs referred to in this work covers 75.3% of the exotic genome (Zahn et al., 2020). Additional information regarding the elite cultivars is given in Table S2.

Flag leaves were collected at two developmental stages: shooting and maturity (BBCH39 and BBCH59, respectively; BBCH-scale)
| Peak | Compound                          | Abbr. | Molecular formula | Monoisotopic mass | Theoretical $[\text{M + H}]^+$ | Bucket (Rt; m/z) | Combined ions | Rt (min) | MS intensity |
|------|----------------------------------|-------|-------------------|-------------------|---------------------------------|-----------------|---------------|----------|--------------|
| 1    | Protocatechuic acid              | PCA   | C$_7$H$_6$O$_4$   | 154.0266          | 155.0344                        | 2.02; 155.0336  | $[\text{M + H}]^+$ $[\text{M + H - H$_2$O}]^+$ | 2.02 (0.44) | 82,298 (21.85) |
| 2    | p-Hydroxybenzoic acid            | PHB   | C$_7$H$_6$O$_3$   | 138.0317          | 139.0395                        | 2.70; 139.0387  | $[\text{M + H}]^+$                                  | 2.70 (0.33) | 89,941 (19.23) |
| 3    | Chlorogenic acid                 | CGA   | C$_{12}$H$_{18}$O$_9$ | 354.0951         | 355.1029                        | 3.22; 355.1027  | $[\text{M + H}]^+$ $[\text{M + Na}]^+$ $[\text{M + H - C$_7$H$_12$O$_3$}]^+$ | 3.22 (0.20) | 188,774 (24.84) |
| 4    | Dihydromyricetin                 | MYR   | C$_{13}$H$_{12}$O$_8$ | 320.0532         | 321.0610                        | 4.05; 321.0608  | $[\text{M + H}]^+$                                  | 4.05 (0.14) | 308,118 (17.28) |
| 5    | Rutin (Quercetin 3-O-rutinoside) | RUT   | C$_{22}$H$_{30}$O$_{16}$ | 610.1534         | 611.1612                        | 4.93; 611.1618  | $[\text{M + H}]^+$ $[\text{M + Na}]^+$ $[\text{M + H - C$_{12}$H$_2$O$_4$}]^+$ | 4.93 (0.11) | 208,384 (34.24) |
| 5    | Rutin (Aglycone ion)             | RUT$^a$ | C$_{22}$H$_{30}$O$_{16}$ | 610.1534         | 611.1612                        | 4.93; 303.0501  | $[\text{M + H}]^+$ $[\text{M + Na}]^+$ $[\text{M + H - C$_{12}$H$_2$O$_4$}]^+$ | 4.93 (0.11) | 21,329 (14.94) |
| 6    | Naringin (Naringenin 7-O- rhamnoglucoside) | NIN$^a$ | C$_{22}$H$_{32}$O$_{14}$ | 580.1792         | 581.1870                        | 6.13; 581.1875  | $[\text{M + H}]^+$ $[\text{M + Na}]^+$ $[\text{M + H - C$_{12}$H$_2$O$_4$}]^+$ $[\text{M + H - C$_6$H$_10$O$_4$}]^+$ $[\text{M + H - C$_6$H$_10$O$_5$}]^+$ | 6.13 (0.08) | 149,965 (34.98) |
| 6    | Naringin (Aglycone ion)          | NIN$^a$ | C$_{22}$H$_{32}$O$_{14}$ | 580.1792         | 581.1870                        | 6.13; 273.0760  | $[\text{M + H}]^+$ $[\text{M + Na}]^+$ $[\text{M + H - C$_{12}$H$_2$O$_4$}]^+$ $[\text{M + H - C$_6$H$_10$O$_4$}]^+$ $[\text{M + H - C$_6$H$_10$O$_5$}]^+$ | 6.13 (0.08) | 285,136 (16.71) |
| 7    | Quercetin                        | QUE   | C$_{15}$H$_{10}$O$_7$ | 302.0427         | 303.0505                        | 7.05; 303.0502  | $[\text{M + H}]^+$                                  | 7.05 (0.07) | 70,628 (15.73) |
| 8    | Naringenin                       | NEN   | C$_{15}$H$_{12}$O$_5$ | 272.0685         | 273.0763                        | 8.14; 273.0759  | $[\text{M + H}]^+$                                  | 8.14 (0.06) | 195,747 (15.96) |

Note: Description of the eight soluble phenolic compounds comprising the standard mix analyzed by RP-LC-PDA-ESI-HR-QTOF-MS. The corresponding buckets (Rt; m/z pairs) are indicated together with the adduct and fragment ions combined during data processing. The average Rt and MS intensity for 100 injections along more than 2500 LC–MS runs are indicated for every compound (CV-Coefficient of Variation—in parenthesis). Peak numbers are according to Figure 2A.

$^a$Denotes aglycone ions.
2.3.2 | Sunflower

Sunflower accessions (H. annuus L.) were grown in 2017 in the field at the IPK-Gatersleben, Germany (51° 49' 19.74" N, 11° 17' 11.8" E; black clayey loam soil; annual temperature: 9°C; annual precipitation: 486 mm). The fully developed second leaves (starting from the flower basis) of three different plants were harvested per accession and immediately frozen in liquid nitrogen for further analysis. Five accessions were analyzed: A1 (HEL166, Georgia), A2 (HEL159, Canada), A3 (HEL208, Georgia), A4 (HEL697, USA), A5 (HEL237, Bulgaria).

2.3.3 | Arabidopsis thaliana

Leaves were harvested from five-week-old Arabidopsis Col-0 plants grown under in vitro nitrogen (N) sufficiency and deficiency conditions (Józefowicz et al., 2020). Six plant replicates per condition were collected and immediately frozen in liquid nitrogen for further analysis.

2.4 | Untargeted phenylpropanoid metabolotyping workflow

The high-throughput untargeted phenylpropanoid metabolotyping method was implemented based on previous protocols (Brauch et al., 2018) as a starting point. The workflow comprises six major stages (Figure 1).

2.4.1 | Step 1: Metabolite extraction

Freeze-dried or liquid nitrogen-frozen leaf material, contained in 20 mL vials (HDPE; Zinsser Analytic GmbH, DE), were added with two-to-four steel beads (Ø 8 mm; Mühlmeier GmbH&Co. KG). Grinding was carried out in a Retsch mixer mill MM 400 (Retsch GmbH) to the RP-UPLC-PDA system. Phenylpropanoids were further analyzed via ESI-UHR-QTOF (Electrospray Ionization-Ultra-High Resolution-Quadrupole Time Of Flight) MS by coupling a maXis Impact ESI-QTOF MS (Bruker Daltonik GmbH) to the RP-UPLC-PDA system.

2.4.2 | Steps 2 and 3: RP-UPLC-PDA Analysis

Phenylpropanoids were analyzed by RP-UPLC-PDA (Reversed Phase Ultra Performance LC separation coupled to PhotoDiode Array detection) using an Acquity UPLC system (Waters), equipped with an Acquity UPLC PDA eX explorer detector (Waters). Before their analysis, 80 μL aliquots from the methanolic extracts were mixed with 20 μL of 0.5% (v/v) formic acid, incubated overnight (−20°C), and centrifuged (22,500g, 4°C, 10 min) to remove precipitates. Samples were randomized before their analysis. Sample injection was performed using a PLNO (Partial-Loop with Needle-Overfill) injection mode. The compound separation was carried out with different RP-LC methods depending on the sample's chemistry as described in Table S4. PDA-detection was performed in a range between 210 and 800 nm, at a resolution of 1.2 nm, and a sampling rate of 20 points s⁻¹.

2.4.3 | Step 4: MS Acquisition

Phenylpropanoids were further analyzed via ESI-UHR-QTOF (Electrospray Ionization-Ultra-High Resolution-Quadrupole Time Of Flight) MS by coupling a maXis Impact ESI-QTOF MS (Bruker Daltonik GmbH) to the RP-UPLC-PDA system.

Owing to the larger molecular weight of anthocyanins, which were present in Arabidopsis samples, the following MS1 parameters in positive mode were modified based on previous protocols (Oertel et al., 2017): 50–1500 m/z; capillary voltage: 4 kV; nebulizer: 3 bar; dry gas: 8 L min⁻¹; dry temperature: 200°C; hexapole RF (Ratio Frequency) voltage: 60 Vpp (V peak-to-peak); funnel 1 RF: 300 Vpp; funnel 2 RF: 300 Vpp; pre-pulse storage time: 8 μs; transfer time: 60 μs; low mass: 40 m/z; collision cell RF: 800 Vpp; collision energy: 3 eV. For annotation purposes, MS/MS was performed in auto MS/MS mode using CID (Collission-Induced Dissociation) with the following settings: absolute area threshold: 5000 counts; exclusion activation: 15 spectra; exclusion release: 60 s; collision energy values (z = 1, 2, 3; isolation mass = 500; width = 8): 15, 10, 5 eV; collision energy values (z = 1, 2, 3; isolation mass = 1000; width = 10): 25, 20, 15 eV. Similar parameters as those indicated above were employed for the MS1 analysis in negative mode, except the following: capillary voltage: 3.5 kV; hexapole RF voltage: 100 Vpp.

The system was calibrated before each run with 10 mM sodium formate (water: isopropanol 1:1 v/v) at an infusion flow enable recovering 95% of the total soluble UV-absorbing compounds (Figure S1).
rate of 0.12 mL h\(^{-1}\), using an enhanced quadratic calibration mode. The Compass HyStar 3.2 SR2 software (Bruker Daltonik GmbH) was used to operate and coordinate LC-PDA-MS data acquisition.

2.4.4 | Steps 5 and 6: Data processing and analysis

Simultaneous processing of hundreds of raw LC-MS data files was performed using MetaboScape 5.0 (Bruker Daltonik GmbH). The
software applies a time-aligned region complete extraction algorithm (T-Rex 3D) for mass calibration, non-linear retention time alignment, feature extraction, de-isotoping, de-adducting, and ion combination (i.e., isotopes, charge states, adducts, or fragments) into single features. The processing of raw data signals results in a single-aligned matrix (bucket table) comprised of the identified “buckets” (retention time-m/z pairs) and their abundance, in terms of either signal intensity or area, across the analyzed samples. Data processing was carried out following the following parameters: intensity threshold: $10^7$–$10^9$ counts; minimum peak length: 8 spectra; minimum peak length (recursive): 7 spectra; mass and Rt (Retention time) ranges: defined according to the LC-MS methods; mass recalibration: automatic; EIC (Extracted Ion Chromatogram) correlation for ion deconvolution: 0.8; primary and seed ions: [M + H]$^+$, [M + Na]$^+$, [M + K]$^+$, [M - H]$^-$; common ions: [M + H-H$_2$O]$^+$, [M + H-NH$_3$]$^+$, [2 M + H]$^+$, [2 M + Na]$^+$, [M + H-C$_6$H$_5$O$_2$]$^+$, [M + H-C$_6$H$_5$O$_4$]$^+$, [M + H-H$_2$O]$^+$, [M + H-C$_6$H$_5$O$_2$]$^+$, [M + H-C$_6$H$_5$O$_4$]$^+$, [M + H-H$_2$O]$^+$, [M - H - CO$_2$H$_2$]$^-$, [M - H - NH$_3$]$^-$, [M - H - CO$_2$H$_2$]$^-$, [M - H - NH$_3$]$^-$.

Data transformation (e.g., normalization, scaling), statistical analysis, and compound identification were also performed with MetaboScape 5.0. The buckets derived from blank injections were removed by using the bucket exclusion feature, which was parameterized to exclude all buckets of which the intensity in biological samples did not exceed the intensity in blank samples by at least factor three. Using this software, PCA (Principal Component Analyses) were carried out with these settings: scaling: Pareto; cross-validation mode: 10%; minimum explained variance: 98%. Compound annotation was performed by manual inspection of MS/MS fragmentation patterns and supported by features from the MetaboScape software (e.g., SmartFormula and Spectral Library with m/z tolerance of 2–5 mDa). Manual inspection of LC-PDA-MS data was performed using the Compass DataAnalysis 4.4 SR1 package (Bruker Daltonik GmbH).

3 | RESULTS

3.1 | Workflow performance across multiple runs: Using a phenolics standard mix

As an initial verification of our approach, a set of eight metabolites comprising soluble single and polyphenolic compounds of different polarities and masses was defined to assess the workflow performance (Table 1). For this purpose, we evaluated a dataset of 100 injections of the standard mix distributed along more than 2500 LC-MS runs that were performed not in a single, but along five different batches. The workflow proved robust allowing to process the 100 LC-MS chromatograms into a single data matrix, showing a maximum Rt drift of 1.2 s (maximum CV [coefficient of variation] of 0.44%) in the MS signals (Figure 2A, Table 1). This enabled the proper identification of the standard features and the combination of multiple ions into single buckets, thus reducing feature replication. The MS intensities displayed changes within and along the batches depending on compound chemistry (Figure 2B, Table 1). Flavonoid glycosides (RUT, NIN) showed the highest variations (35% CV), followed by phenolic acids (PCA, PHB) and their esters (CGA), while flavonoid aglycones (QUE, NEN) displayed the lowest (approx. 15% CV). Moreover, different variation patterns were observed within the ions combined into single metabolic features. Contrary to the molecular ions [M + H]$^+$ of the flavonoid glycosides, the aglycone ions derived from their in-source fragmentation (RUT$, NIN^+$) showed lower susceptibility to batch variability (approx. 15% CV). These results underscore the relevance of using internal and external quality controls to monitor systematic variations in the workflow by using not a single, but multiple metabolites that mimic the composition of the sample to analyze. Using additional standards for positional C-glucosylflavone isomers frequently found in cereals, the workflow proved to discriminate them based on their chromatographic separation (Figure S2).

To evaluate the feasibility of using these compounds as an internal standard, barley leaf extracts were spiked with the standard mix (Figure 2C). As we have not yet detected these compounds in our barley extracts, the standard mix proved useful as an internal and external standard for the high-throughput LC-MS analysis of barley phenylpropanoids.

3.2 | Phenylpropanoid profiles of the elite barley cultivars Barke and Scarlett

Two German spring barley cultivars, cv. Barke and Scarlett, both of which are parents of different mapping populations, were chosen to further evaluate the workflow. Firstly, the flag leaves of both genotypes grown in field conditions were analyzed by LC-PDA-MS/MS, and the major soluble UV-absorbing compounds were annotated based on their absorption and MS-fragmentation spectra (Figure 3, Table S5). The phenylpropanoid profiles were dominated by glycosylation and acylation derivatives of three flavone 6-C-glucosides: isovitexin, isoorientin, and isoscoparin, where the latter two comprise an additional 3'-hydroxyl and a 3'-hydroxymethyl group in their aglycone moieties, respectively, when compared to isovitexin (Figure 3). The workflow discriminated different positional O-glycosylation isomers of these C-glucosylflavones (e.g., peaks 6, 9, 11, 15, 16, 17). Hydroxycinnamic acid derivatives (peaks 1, 4, 5) and a putative flavonol C-hexoside (peak 2) were also detected in lower proportions. The differences in O-glycosylation patterns of the 6-C-glucosylflavones, were responsible for the contrasting phenylpropanoid profiles observed between both genotypes: 7-O-glycosyl-6-C-glucosylflavones were predominant in cv. Barke, whereas cv. Scarlett was dominated by 2”-O-glycosyl-6-C-glucosylflavones (Table S5). The C-pentosides of these 6-C-glucosylflavones (peaks 8, 12, 14) were common to both cultivars. Making use of their contrasting phenylpropanoid profiles, these genotypes were employed to evaluate the high-throughput performance of the metabotyping workflow with plant extracts.

3.3 | High-throughput phenylpropanoid metabotyping of contrasting barley genotypes

Next to data acquisition, a major bottleneck for high-throughput untargeted LC-MS-based analyses is large data processing without compromising the identification of biologically relevant features. An
LC–MS dataset was generated from 205 injections comprising repeated analyses of extracts from the two contrasting cultivars Scarlett and Barke, as well as from a mixed sample of both. First, the classical auto MS/MS fragmentation strategy (i.e., alternating between the acquisition of low-energy mass precursor ion spectra and high-energy fragment ion spectra) was tested in positive polarity. Using a
low MS intensity threshold (1500 spectral counts), the simultaneous processing of 205 LC–MS/MS chromatograms resulted in a single data matrix of 35,231 buckets across all the samples. A PCA of this dataset proved the capability to discriminate barley genotypes based on their phenylpropanoid metabotypes without previous knowledge of the underlying features (Figure 4A). When considering the annotated compounds (Table S5), the isoorientin and isovitexin moieties that displayed contrasting O-glycosylation patterns, were the most influential to distinguish both cultivars (Figure 4A, loadings plot). The workflow not only distinguished both cultivars but also discriminated within the samples according to their different abundances of common metabolites. This was demonstrated by the clear clustering of the combined extracts from both genotypes (mix) into a third group (Figure 4A).

To further increase the number of spectra per chromatographic peak and quantitation accuracy, an alternative approach to auto MS/MS was evaluated. Since the possibility of importing and matching MS/MS data from selected measurements rather than from the whole dataset is enabled by the MetaboScape software, an MS1-based approach was tested with the same array of samples. Using MS1-based acquisition and a low MS-intensity threshold, 8720 additional buckets were detected in comparison to the MS/MS approach. Together with an increased number of scans per feature, this resulted in a tighter sample clustering, improved metabotype-dependent genotype discrimination (Figure 4B), and reduced relative quantitation variability (Figure 4D). Owing to the increased spectra acquired per feature, the MS1 dataset was reprocessed using a high-intensity threshold (10,000 spectral counts) to ease data processing and analysis. Compared to the low-intensity threshold, a decreased number of buckets (30-fold) were identified without affecting the detection and quantitation of those relevant for genotype discrimination (Figures 4C,D). The most influential features (e.g., isovitexin and isoorientin O-glycosides) still contributed to sample discrimination across the first principal component (Figure 4C, loadings plot), which explained almost 50% of the dataset variance (Figure 4C, scores plot).

Phenylpropanoids are also detected in negative ionization mode (Olmo-García et al., 2018; Piasecka et al., 2015). To evaluate the effect of additionally using data acquired in negative mode, flag leaf extracts from cv. Barke and Scarlett were analyzed by LC–MS1 in both, positive and negative polarities. When comparing the number of features detected in each polarity, the positive mode enabled the detection of more features (4.2-fold) than those found in negative mode (Figure S3A,B). After merging the data from both polarities, a dataset comprising only 13% more features than those detected in positive mode was generated. The PCA analysis of the three datasets (i.e., positive, negative, and merged polarities) showed that independent from the ionization mode, the most influential features were still
the C-glucoflavone derivatives with contrasting O-glycosylation patterns (Figure S3). Altogether, using a negative ionization mode for the high-throughput profiling of phenylpropanoids in barley leaves would not change, nor significantly improve the outcomes from an LC–MS workflow in positive mode, whereas the number of LC–MS runs would be duplicated.
Overall, these results show that an LC–MS1-based workflow in positive ionization mode using a high-intensity MS-threshold, enabled the high-throughput analysis of barley extracts without compromising their discrimination. The extracts were distinguished according to their phenylpropanoid patterns, where the contribution of the most contrasting soluble UV-absorbing compounds was evident.
3.4 Phenylpropanoid metabotyping of complex datasets: Barley mapping populations

The performance of the workflow was evaluated with a complex dataset comprising more than two barley genotypes at different developmental stages. We analyzed the flag leaves of field-grown IL from the barley S42IL mapping population (Honsdorf et al., 2017; Zahn et al., 2020), together with seven additional elite cultivars. The processing of the LC–MS1 data using a high-intensity threshold resulted in a matrix of 1713 buckets across 241 samples.

Two major factors challenged the workflow performance. First, the genetic closeness of the IL to the parent cv. Scarlett tested the method’s capability to discriminate contrasting genotypes within lines displaying similar phenylpropanoid patterns. Second, the different developmental stages tested the workflow’s capacity to distinguish within development-dependent effects. To assess the advantages of using an LC–MS-based metabotype compared to the information commonly retrieved from an LC-PDA analysis, we firstly analyzed a subset comprising 11 buckets that represented the major soluble UV-absorbing phenylpropanoids in barley flag leaves. The poor sample clustering in the PCA scores plot indicated that the MS1-based intensities of the 11 phenolic compounds readily detectable by LC-PDA analysis could not fully represent the underlying dataset complexity (Figure 5A). In contrast, when using the complete LC–MS-based metabotype, two clusters discriminating between the developmental stages were clearly defined across the first principal component. Additional sample separation across the second principal component was observed based on the contrasting 7- and 2′-O-glycosylation of the C-glucoflavones (Figure 5B). Hence, the phenylpropanoid metabolotypes not only discriminated within contrasting flavonoid patterns but also reflected the complexity of the developmental changes occurring in flag leaves. The relative quantitation (Figure 5C, Table S3) of contrasting features from the loadings plot (Figure 5B, right panel) suggested that the developmental changes might be led by phenylpropanoids different from the major O-glycosylated 6-C-glucoflavone derivatives, proving the potential of untargeted metabotyping to identify novel features of interest.

In a previous study (Brauch et al., 2018), the phenylpropanoid profiling of young greenhouse-grown seedlings from the S42IL population using an LC-PDA approach led to the identification of IL (S42IL-101, 177, 178) devoid of the isovitexin 2′′-O-glicoside. This enabled the identification of a putative glycosyltransferase involved in flavonoid metabolism (Brauch et al., 2018). The identification of the same lines using the LC–MS workflow described here would validate the feasibility of using the buckets’ MS intensities as metabolic quantitative traits. For this purpose, we analyzed the distribution of isovitexin 2′′-O-glicoside and additional O-glycosylated 6-C-glucoflavones across field-grown lines of the S42IL population in stage A (shooting), at which these flavones were detected at high abundance (Figure 5C, Table S3). The flag leaves of most lines displayed a similar abundance of isovitexin and isoorientin 2′′-O-hexosides to those shown by the parental cv. Scarlett. Only a few lines showed decreased average amounts, where the lowest were displayed by the S42IL-101 genotype (Figure 5D, Table S3). This is in line with our previous findings (Brauch et al., 2018); the two lines S42IL-177 and 178 were not identified, as these were absent in the panel analyzed in this work. When compared to cv. Barke as a negative control for the presence of the isovitexin 2′′-O-glucoside (Figure 5D), a higher average amount was still detected in the S42IL-101 flag leaves. This was due to the clear presence of this compound in some biological replicates (four out of 10) rather than by an overall higher abundance, as confirmed by manual inspection of the EIC from individual injections. Similar to cv. Barke, the five additional elite spring cultivars analyzed in this work (cv. Grace, Propino, Quench, RGT Planet, Salome) did not show 2′′-O-glycosylated 6-C-glucoflavones in their profiles (Figure 5D, Table S3).

These results demonstrate the applicability of the workflow for the high-throughput phenylpropanoid metabotyping of complex barley populations and the feasibility of using the detected buckets for further mQTL analyses.

3.5 Workflow application to other plant species

To show the versatility of this workflow, examples are given for its application in dicotyledonous species known to display phenylpropanoid patterns that drastically differ from those of monocots such as barley: the model plant Arabidopsis thaliana, whose leaf phenylpropanoids mostly comprise flavonol derivatives (Tohge et al., 2005); and the crop plant Helianthus annuus L. (sunflower) that displays chlorogenic acid derivatives as its major soluble phenolics while flavonoids, mainly methoxylated flavones, are present in trace amounts (Stelzer et al., 2019). To test the performance of the MS workflow, it was necessary to adjust the chromatographic conditions to the contrasting phenylpropanoid chemistries of these species (Table S4).

The leaves of five sunflower accessions from the IPK genebank grown in field conditions were analyzed (A1 to A5, described in Materials and Methods). Despite their similar LC-PDA profiles, mainly dominated by CGA (3-O-cafeoyl quinic acid) and 3,5-di-O-cafeoyl quinic acid (Figure 6A), the accessions differed in their phenylpropanoid metabolotypes as indicated in the PCA plot (Figure 6B). Further analysis showed that these differences did not rely on the most abundant phenolics (e.g., CGA) but upon novel features. This is illustrated by an unknown UV-absorbing compound (4.16 min; m/z 251.16), which was completely absent in accession A1 while displaying differential amounts in the remaining accessions (Figure 6B).

The application of this workflow to the model Arabidopsis was exemplified by comparing the leaf metabolotypes of plants treated under N-sufficient and N-deficient conditions (Józefowicz et al., 2020). The well-known induction of phenylpropanoid accumulation upon N deficiency was evident when comparing the LC-PDA chromatograms (Figure 6C). This was supported by the differential
sample clustering in a treatment-dependent manner (Figure 6D) that was highly influenced by the levels of kaempferol and cyanidin derivatives. The metabotyping profiles revealed that contrary to the major UV-absorbing compounds, other semi-polar metabolites might decrease upon N deficiency. This is exemplified by the quantitation of an unknown feature (0.47 min; m/z 476.86).
DISCUSSION

Natural variation represents a valuable source to address phenylpropanoid diversity in cereals and to elucidate its genetic basis but poses a challenge by demanding large-scale analyses. Here, a coupled LC-PDA-MS workflow for complex phenylpropanoid profiling across hundreds of samples has been established and verified. MS data were acquired in MS1 positive mode, facilitating further data processing and analysis through the proprietary software MetaboScape. Robustness of the analytical performance as well as of the data analysis were tested using repeated injections of (1) a standard mix, and (2) extracts from selected barley genotypes. The standard mix analysis uncovered the dependence of batch-to-batch variations on the compound’s chemistry. The analysis of barley extracts in large datasets proved the workflow’s capability to discriminate within genotypes, highlighting the role of glycosylation in barley phenylpropanoid diversity. Analysis of the barley S42-IL mapping population validated the utility of using the metabolotypes’ MS features for metabolic quantitative trait purposes. The metabolotype complexity also reflected development-dependent changes. After adjustment of the chromatographic conditions, the workflow was successfully applied for two other plant species whose phenylpropanoid profiles differ from those of cereals: sunflower and the model plant Arabidopsis. This approach enables high-throughput phenylpropanoid metabolotyping and its application to large-scale natural variation studies.

The interest in studying the diversity of barley phenylpropanoid metabolism is not recent. Early studies profiling nearly 1500 barley varieties through two-dimensional TLC, already indicated the complexity of barley flavonoid patterns and their remarkable variability across different cultivars (Fröst et al., 1975). A total of 27 flavonoids enabled their classification into five chemical races, each of them corresponding to a different geographical distribution. It was proposed that race-specific chemical differences relied on biosynthetic flavonoid decorations commonly affecting more than one compound (e.g., 3’-O-methylation, glycosylation) (Fröst et al., 1977). In recent years, 152 phenylpropanoids were identified in nine barley varieties by in-depth analysis using MS and NMR (Nuclear Magnetic Resonance). The study confirmed phenylpropanoid glycosylation diversity in barley, but a UV profile-based chemotaxonomic analysis did not discriminate cultivars in a geographic-wise manner (Piasecka et al., 2015). The barley metabolotypes generated in this work support indications from prior studies. First, our results confirm that the phenylpropanoid profiles of barley leaves are dominated by glycosylation and acylation derivatives of flavone glycosides, mainly from the 6-C-glucoflavones isovitexin, isoorientin, and isoscoparin to a lower extent. Second, the major glycoflavone in barley leaves is tissue-dependent: as shown previously (Fröst et al., 1977) as well as in this work, isoorientin derivatives lead the profiles in flag leaves, whereas isovitexin derivatives are dominant in the leaves of young seedlings (Brauch et al., 2018; Kaspar et al., 2010; Reuber et al., 1996). Third, the contrasting metabolotypes from cv. Barke and Scarlett support the O-glycosylation of common C-glycoflavones as a major source of diversity in barley phenylpropanoids. This points toward specific glycosyltransferases as major determinants of barley flavonoid diversity. This is in line with the identification of a candidate glycosyltransferase involved in the synthesis of the isovitexin 2’-O-glucoside (Brauch et al., 2018), and the proven role of glucosyltransferases in the natural variation of rice flavone accumulation (Peng et al., 2017).

Aiming to study the genetic basis underlying barley phenylpropanoid diversity, we extended our prior LC-PDA-based approach (Brauch et al., 2018) to the MS approach here described. The multidimensional data from LC–MS workflows not only provide great advantages owing to the high detection sensitivity and coverage but also pose a challenge for complex data processing and analysis. We challenged the MetaboScape software which has been successfully employed for studies on specialized plant metabolites (Olmo-García et al., 2018; Villette et al., 2018). With the workflow here described, 200 LC–MS runs were processed in periods of one to a maximum of two hours. Initial work using this workflow with mapping populations is enabling to process more than 1600 LC–MS runs in a single dataset and thus, to overcome limitations of large LC–MS data processing.

An MS1 acquisition mode instead of the traditional auto MS/MS metabolomics approach was implemented in this workflow. This improved the quality of the acquired MS data in terms of increased detected buckets and scans per feature, resulting in a more robust feature quantitation and metabolotype-based sample discrimination. To increase the annotation likelihood enabled by spectral matching, the MS/MS data from single representative measurements (e.g., pool samples) can be merged to the MS1 features using this workflow. It must be highlighted that compound annotation, particularly for plant specialized metabolites, remains challenging. Dedicated libraries (e.g., ReSpect, Weizmass) (Sawada et al., 2012; Shahaf et al., 2016) and/or tailored computational tools such as FlavonoidSearch (Akimoto et al., 2017) and FlavonQ (Zhang et al., 2017) can support phenylpropanoid annotation. Using exclusively a positive ionization mode, the workflow from this study already generated biologically relevant phenylpropanoid metabolotypes. Merging data of different polarities did not significantly improve the outcomes from the workflow in positive mode. However, assessing the effect of different ionization modes is always recommended, since the outcomes strongly rely on the metabolites’ chemistry.

The metabolotyping workflow still faces limitations for phenylpropanoid analysis, which are related to common challenges for metabolomics feature processing; (1) metabolite over-representation in different buckets; (2) combination of different metabolites into single buckets. The first issue is often a consequence of the in-source fragmentation to which O-glycosylation moieties are particularly prone (Akimoto et al., 2017; Kachlicki et al., 2016). To cope with this, we include the early neutral losses of O-glycosyl moieties in the ion configuration rules for bucket combination (e.g., $[M + H-C_6H_2O_3]^{+}$ and $[M + H-C_6H_2O_2]^{+}$, for hexosyl and deoxyhexosyl moieties, respectively). The second issue is related to the combination of close-eluting phenylpropanoid glycosylation isomers into single buckets. The adjustment of Rt ranges for feature...
alignment and grouping in the processing algorithm enable coping with this complication. Undoubtedly, the continuous evolution of computational tools and the integration of novel technologies in LC–MS workflows, is opening new possibilities in phytochemical analysis. Trapped ion mobility spectrometry coupled to LC–MS, has already proven to discriminate within multiple groups of flavonoid isobaric and isomeric compounds that may not be resolved through their chromatographic separation. The performance of this approach was demonstrated by using a mixture of seven monohydroxyflavones (2′-, 3′-, 4′-, 3-, 5-, 6-, 7-hydroxyflavone), where 90% of the combinatorial possibilities were separated; additional sets of flavones and flavanones with different hydroxy- and/or methoxylation patterns were evaluated (Schroeder et al., 2019).

Elucidating the genetics underlying phenylpropanoid metabolism demands the integration of genetic association studies to in-depth metabolomic profiling. We tested and demonstrated the feasibility of the metabolotyping workflow for mQTL purposes by profiling a subset of the barley S42IL mapping population, which was previously analyzed using an LC-PDA approach (Brauch et al., 2018). In both strategies, the same line was identified (S42IL-101) for displaying low or negligible levels ofisorvitexin 2′-O-glucoside. The presence of low levels of this compound in the flag leaves of some S42IL-101 plants (four out of 10), compared to its absence in the second leaves of young S42IL-101 seedlings (Brauch et al., 2018), could be attributed to different factors: the residual heterozygosity of this line (2.5%) (Honsdorf, 2017) and/or the metatypotype susceptibility to the variations inherent to field conditions (e.g., differences in nutrient availability). The generated metatypotypes reflected not only the genotypic complexity of mapping populations but also their development-dependent changes. Using MS-based metabolite mapping is a strategy that has already provided novel insights into the specialized metabolism of cereals (rice, maize, and quingke), mainly through wide-targeted metabolomics based on multiple reaction monitoring (Chen et al., 2013; Peng et al., 2017; Wen et al., 2014; Zeng et al., 2020).

This is a highly sensitive and accurate MS method that is nevertheless, restricted to a predefined set of metabolites based on the prior construction of an MS/MS spectral tag library (Chen et al., 2013). In the untargeted phenylpropanoid metabotyping workflow described here, all the detectable features are measured. This readily enables to assess their presence and influence in the dataset diversity, without prior construction of a library or compound annotation. Hence, unknown features might be used as quantitative traits for their association with specific genetic loci (Wu et al., 2018). As already shown for human serum samples (Krumsieck et al., 2012), this approach could support predicting the identity of unknown metabolites, which is a common bottleneck in phytochemistry analyses.

The genetic diversity represented by the wild and domesticated barley accessions maintained at the German national Genebank hosted by the IPK, together with the continuous development of genome-sequencing resources, are enabling the accurate assessment of genetic variation in these highly diverse germplasm collections (Jayakodi et al., 2020; Milner et al., 2019). This is complemented by the availability of mapping populations with a high genetic resolution, such as the Morex X Barke Recombinant Inbred Lines (Mascher et al., 2013) and the ‘Halle Exotic barley 25’ Nested Association Mapping (HEB-25 NAM) population (Maurer et al., 2015); these enable investigating specific genomic regions and their allelic variation. Using these resources, known and novel loci underlying phenotypic traits have been identified and allocated to different barley gene pools. The application of the metabolotyping workflow to these resources provides opportunities to investigate the diversity and specific roles played by specialized metabolites, as well as to identify the underlying genes. Moreover, multiple subjects can be addressed: Do different genetic pools display contrasting metatypotypes? Are the metatypotypes responsible for specific phenotypic traits? Can novel metabolites and their underlying genes be retrieved from exotic genotypes? The workflow described here allows exploring additional resources available at the IPK Genebank, as exemplified here with sunflower. The method is also applicable to the model species Arabidopsis. Using metabolotyping approaches in natural genetic variation-based studies and their further integration to other “omics”-based studies (e.g., transcriptomics, proteomics), will provide exciting insights into the phenylpropanoid metabolism of crop plants. By addressing current challenges in the field of specialized metabolites, the workflow here described can be readily applied to study in a high-throughput fashion the relevance underlying the natural diversity of phenylpropanoids in plants, particularly barley.

ACKNOWLEDGMENTS
Funding for our research by the German Federal Ministry of Research and Education (BMBF) through the IPAS grant BARLEY-DIVERSITY (FKZ 031A3528) is acknowledged. We thank Prof. Dr. Klaus Pillen (Martin Luther University, Halle-Wittenberg, DE) for providing the barley flag leaf samples. We thank the excellent technical assistance of Annegret Wolf, Elena Brueckner, Franziska Hartmann, and Sebastian Fricke. We also gratefully acknowledge Dr. Stefanie Döll for her support during the implementation of this workflow.

CONFLICT OF INTEREST
Nikolas Kessler is an employee of Bruker Daltonik GmbH which manufactures and sells the mass spectrometer and software used in this study.

AUTHOR CONTRIBUTIONS
Conceptualization: Adriana Garibay-Hernández, Hans-Peter Mock; Arabidopsis metabolite extraction and data acquisition: Adriana Garibay-Hernández, Anna Maria Jozefowicz; Sunflower metabolite extraction and data acquisition: Adriana Garibay-Hernández, Gözde Merve Türksoy; Genebank resources: Ulrike Lohwasser; Data processing and analysis: Adriana Garibay-Hernández, Nikolas Kessler; Adriana Garibay-Hernández and Hans-Peter Mock wrote the manuscript. All authors reviewed and approved the final manuscript.

DATA AVAILABILITY STATEMENT
The data supporting this study are openly available through the e!DAL electronic data archive library (Arend et al., 2014) at: https://doi.org/10.5447/ipk/2021/11.
Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T. et al. (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature, 544, 427–433.

Maurer, A., Draba, V., Jiang, Y., Schnaitmann, F., Sharma, R., Schumann, E. et al. (2015) Modelling the genetic architecture of flowering time control in barley through nested association mapping. BMC Genomics, 16, 290.

Meldgaard, M. (1992) Expression of chalcone synthase, dihydroflavonol reductase, and flavanone-3-hydroxylase in mutants of barley deficient in anthocyanin and proanthocyanidin biosynthesis. Theoretical and Applied Genetics, 83, 695–706.

Milner, S.G., Jost, M., Taketa, S., Rey Mazón, E.H., Oppermann, M., Weise, S. et al. (2019) Genebank genomics highlights the diversity of a global barley collection. Nature Genetics, 51, 319–326.

NDong, C., Anzellotti, D., Ibrahim, R.K., Huner, N.P.A. & Sarhan, F. (2003) Daphnetin methylation by a novel O-methyltransferase is associated with cold acclimation and photosystem II excitation pressure in rye. The Journal of Biological Chemistry, 278, 6854–6861.

Neelam, Khattak, A. & Sharma, K.K. (2020) Phenylpropanoids and its derivatives: biological activities and its role in food, pharmaceutical and cosmetic industries. Critical Reviews in Food Science and Nutrition, 60, 2655–2675.

Newton, A.C., Valentine, T.A., McKenzie, B.M., George, T.S., Guy, D.C. & Hackett, C.A. (2020) Identifying spring barley cultivars with differential response to tillage. Agronomy, 10, 686.

Norbaek, R., Brandt, K. & Kondo, T. (2000) Identification of flavone C-glycosides including a new flavonoid chromophore from barley leaves (Hordeum vulgare L.) by improved NMR techniques. Journal of Agricultural and Food Chemistry, 48, 1703–1707.

Oertel, A., Matros, A., Hartmair, A., Arapitsas, P., Dehmer, K.J., Martens, S. et al. (2017) Metabolite profiling of red and blue potatoes revealed cultivar and tissue specific patterns for anthocyanins and other polyphenols. Planta, 246, 281–297.

Olmo-Garcia, L., Kessler, N., Neuweger, H., Wendt, K., Olmo-Peinado, J., Fernández-Gutiérrez, A. et al. (2018) Unravelling the distribution of secondary metabolites in Olea europaea L.: exhaustive characterization of eight olive-tree derived matrices by complementary platforms (LC-ESI/APCI-MS and GC-APCI-MS). Molecules, 23, 2419.

Peng, M., Shahzad, R., Goll, A., Subhan, H., Shen, S.Q., Lei, L. et al. (2017) Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. Nature Communications, 8, 1975.

Petridis, A., Döll, S., Michelmann, L., Bilger, W. & Mock, H.-P. (2016) Arabidopsis thaliana G2-LIKE FLAVONOID REGULATOR and BRASSINOSTEROID ENHANCED EXPRESSION1 are low-temperature regulators of flavonoid accumulation. The New Phytologist, 211, 912–925.

Peukert, M., Weise, S., Röder, M.S. & Matthies, I.E. (2013) Development of SNP markers for genes of the phenylpropanoid pathway and their association to kernel and malting traits in barley. BMC Genetics, 14, 97.

Piascecka, A., Sawikowska, A., Krajewski, P. & Kacperska, A. (2015) Combined mass spectrometric and chromatographic methods for in-depth analysis of phenolic secondary metabolites in barley leaves. Journal of Mass Spectrometry, 50, 513–532.

Piascecka, A., Sawikowska, A., Kuczyńska, A., Ogrodowicz, P., Mikolajczak, K., Krzystkowiak, K. et al. (2017) Drought-related secondary metabolites of barley (Hordeum vulgare L.) leaves and their metabolomic quantitative trait loci. The Plant Journal, 89, 896–913.

Pihlava, J.-M. (2014) Identification of hordatines and other phenolamides in barley (Hordeum vulgare) and beer by UPLC-QTOF-MS. Journal of Cereal Science, 60, 645–652.

Poursarebani, N., Trautewig, C., Melzer, M., Nussbaumer, T., Lundqvist, U., Rutten, T. et al. (2020) COMPOSITUM 1 contributes to the architectural simplification of barley inflorescence via meristem identity signals. Nature Communications, 11, 5138.
Vogt, T. (2010) Phenylpropanoid biosynthesis. *Molecular Plant*, 3, 2–20.

Vreugdenhil, D., Aarts, M. & Koornneef, M. (2005) Exploring natural variation to improve plant nutrient content. In: Broadley, M.R. & White, D. J. (Eds.) *Plant nutritional genomics*. Oxford: Blackwell Publishing, pp. 201–219.

Walkowiak, S., Gao, L.L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J. et al. (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature*, 586, 287–283.

Walka, A., van Esse, G.W., Kirschnner, G., Guo, G., Brünje, A., Finkemeier, I. et al. (2020) An acyl-CoA N-acyltransferase regulates meristem phase change and plant architecture in barley. *Plant Physiology*, 183, 1088–1109.

Weise, S., Scholz, U., Röder, M.S. & Matthies, I.E. (2009) MetaBrew: a comprehensive database of malting quality traits in brewing barley. *Barley Genetics Newsletter*, 39, 1–4.

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.

Westengen, O.T., Lusty, C., Yazbek, M., Amri, A. & Asdal, A. (2020) Safeguarding a global seed heritage from Syria to Svalbard. *Nature Plants*, 6, 1311–1317.

von Wettstein, D. (2007) From analysis of mutants to genetic engineering. *Annual Review of Plant Biology*, 58, 1–19.

Wu, S., Tohge, T., Cuadros-Inostroza, Á., Tong, H., Tenenboim, H., Kooske, R. et al. (2018) Mapping the Arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. *Molecular Plant*, 11, 118–134.

Xu, D., Dhiman, R., Garibay, A., Mock, H.-P., Leister, D. & Kleine, T. (2020) Cellulose defects in the Arabidopsis secondary cell wall promote early chloroplast development. *The Plant Journal*, 101, 156–170.

Yang, W.N., Feng, H., Zhang, X.H., Zhang, J., Doonan, J.H., Batchelor, W.D. et al. (2020) Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular Plant*, 13, 187–214.

Zahn, S., Koblenz, B., Christen, O., Pillen, K. & Maurer, A. (2020) Evaluation of wild barley introgression lines for agronomic traits related to nitrogen fertilization. *Euphytica*, 216, 39.

Zeng, X.Q., Yuan, H.J., Dong, X.K., Peng, M., Jing, X.Y., Xu, Q.J. et al. (2020) Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of qingke. *Molecular Plant*, 13, 112–127.

Zhang, M., Sun, J. & Chen, P. (2017) Development of a comprehensive flavonoid analysis computational tool for ultrahigh-performance liquid chromatography-diode array detection-high-resolution accurate mass-mass spectrometry data. *Analytical Chemistry*, 89, 7388–7397.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

---

**How to cite this article:** Garibay-Hernández, A., Kessler, N., Józefowicz, A.M., Türksoy, G.M., Lohwasser, U., Mock, H.-P. (2021) Untargeted metabotyping to study phenylpropanoid diversity in crop plants. *Physiologia Plantarum*, 1–18. Available from: [https://doi.org/10.1111/ppl.13458](https://doi.org/10.1111/ppl.13458)