Received: 20 May 2020 | Accepted: 8 July 2020

DOI: 10.1002/cche.10323

ORIGINAL RESEARCH ARTICLE

Fingerprinting of wheat protein profiles for improved distinction between wheat cultivars and species

Darina Pronin[1] | **Sabrina Geisslitz**[1,2] | **Andreas Börner**[3] | **Katharina A. Scherf**[1,2]

[1]Leibniz-Institute for Food Systems Biology at the Technical University of Munich, Freising, Germany

[2]Department of Bioactive and Functional Food Chemistry, Institute of Applied Biosciences, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[3]Genebank Department, Leibniz Institute of Plant Genetics and Crop Plant Research, Sealand, Gatersleben, Germany

Correspondence
Katharina A. Scherf, Department of Bioactive and Functional Food Chemistry, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.
Email: katharina.scherf@kit.edu

Funding information
This research was funded by the Leibniz Association within the Leibniz Competition 2015, project number K195/2015 (WHEATSCAN: Unraveling the causes for wheat sensitivities). Additional funding came from the IGF Project of the FEI supported via AiF within the program for promoting the Industrial Collective Research (IGF) of the German Ministry of Economic Affairs and Energy (BMWi), based on a resolution of the German Parliament. Project AiF 18355 N.

Abstract

Background and objectives: Wheat protein composition is commonly characterized by reversed-phase (RP)-HPLC-UV after extraction of albumins/globulins, gliadins (ω5-, ω1,2-, α-, and γ-gliadins), and glutenins (high- and low-molecular-weight glutenin subunits). However, this traditional classification does not consider the individual distribution of peaks, resulting in loss of information on protein fingerprints. We developed a new approach to peak integration and evaluated its suitability to differentiate between wheat cultivars and species.

Findings: Integration events were performed every 20 s, and the relative proportions of the peaks were calculated. We compared the traditional and new integration methods on two sample sets, the first comprising 60 common wheat cultivars from 1891 to 2010 and the second comprising 40 common wheat, spelt, durum wheat, emmer, and einkorn cultivars. The new integration method performed better in differentiating old and modern common wheat cultivars and was also applicable to different wheat species.

Conclusions: Unique cultivars were identified that stood out because of their protein composition. Four samples warrant further research to identify the specific proteins that are responsible for the differences.

Significance and novelty: The new integration allowed us to map the cultivar- and species-specific fingerprints, identify cultivars with exceptional protein composition, and group similar cultivars.

KEYWORDS

breeding, gliadins, glutenins, principal component analysis (PCA), reversed-phase (RP)-HPLC, wheat

1 | INTRODUCTION

Wheat belongs to the *Triticeae* of the *Poaceae* grass family and is subdivided into different species according to ploidy levels. Common wheat, also called bread wheat (*Triticum aestivum L.*), and spelt (*Triticum spelta* L.) are hexaploid species (genome AABBDD). Durum wheat (*Triticum durum* Desf.) and emmer (*Triticum dicoccon* (Schrank) Schübler) are tetraploid species (genome AABB), and einkorn (*Triticum monococcum* L.) is diploid (genome AA) (Delcour et al., 2012). Spelt,

Correction added on 07 August 2020, after first online publication: Projekt DEAL funding statement has been added.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Cereal Chemistry* published by Wiley Periodicals LLC on behalf of Cereals & Grains Association

emmer, and einkorn are hulled and often referred to as ancient wheats, whereas common wheat and durum wheat are free-threshing and designated as modern wheats (Dubcovsky & Dvorak, 2007; Kasarda, 2013). The gluten proteins of common wheat and durum wheat are responsible for the suitability of processing the grains into bread or pasta, respectively. The grains contain 9%–20% of protein, of which about 80% are storage proteins (gluten proteins) and about 20% are metabolic, protective, and structural proteins (Davis, Cain, Peters, Le Tourneau, & McGinnis, 1981; Wieser, Antes, & Seilmeier, 1998). Wheat proteins are typically divided into three fractions: albumins/globulins, gliadins, and glutenins. Albumins and globulins are water- and salt-soluble. Gliadins are predominantly monomeric proteins soluble in aqueous alcohols, whereas glutenins are polymerized by interchain disulfide bonds and only partly soluble in diluted acids and bases or solvents containing detergents and reducing agents (Scherf, Koehler, & Wieser, 2016). Gliadins and glutenins are further subdivided into protein types, and there are three different ways to classify them. The first is based on the amino acid sequence and classification into sulfur (S)-rich, S-poor, and high-molecular-weight (HMW) prolamins (Shewry, Miflin, & Kasarda, 1984). The second classification is based on relative molecular masses (M_r): (a) HMW group ($M_r \approx 67{,}000$–$88{,}000$), (b) medium-molecular-weight (MMW) group ($M_r \approx 40{,}000$–$55{,}000$), and (c) low-molecular-weight (LMW) group ($M_r \approx 30{,}000$–$42{,}000$). The third classification into gluten protein types is the separation of the gliadin fraction into ω5-, ω1,2-, α-, and γ-gliadins and of the glutenin fraction into HMW-glutenin subunits (GS) and LMW-GS based on electrophoretic mobility (Anjum et al., 2007). Each gluten protein type contains different numbers of single proteins, for example, 7 ω-gliadins, 23 α-gliadins, 13 γ-gliadins, 5 HMW-GS, and 22 LMW-GS as identified in the wheat cultivar Butte 86 (Dupont, Vensel, Tanaka, Hurkman, & Altenbach, 2011). However, these numbers and the protein quantities vary depending on the genotype (G), the environment (E), and the G × E interaction (Geisslitz, Longin, Scherf, & Koehler, 2019; Shewry et al., 2010; Ward et al., 2008). Usually, the classification into albumins/globulins and gluten protein types is applied for the qualitative and quantitative characterization of wheat proteins by means of reversed-phase (RP)-HPLC, that can be used to elucidate relationships between protein structure, functionality (e.g., baking performance), and bioactivity (e.g., potential to elicit immune reactions) (Schalk, Lexhaller, Koehler, & Scherf, 2017). However, one disadvantage is the lack of insight into the distribution of individual peaks for particular protein fractions and types. There are only very few reports in the literature that deal with more in-depth characterizations of protein or peptide profiles. One example is the extensive comparison of gliadin peaks by RP-HPLC to identify the effects of temperature and fertilization on the gliadin composition of a winter wheat cultivar (Daniel & Triboi, 2000), or isolation of different gliadins by preparative HPLC (Bin & Peterson, 2016). Other methods to compare differences in wheat genotypes and species include SDS-PAGE (Lagrain, Rombouts, Wieser, Delcour, & Koehler, 2012), a combination of RP-HPLC and SDS-PAGE (DuPont, Vensel, Encarnacao, Chan, & Kasarda, 2004), a combination of RP-HPLC and MS (DuPont, Vensel, Chan, & Kasarda, 2000), and LC-MS/MS (Bromilow et al., 2017; Colgrave et al., 2015; Geisslitz, America, & Scherf, 2020). Due to the complexity of the gluten protein types, it is even more important to provide accurate information about their qualitative and quantitative composition. These insights are crucial to determine changes in protein distribution, distinguish between cultivars and species, and identify protein fingerprints using RP-HPLC patterns.

Therefore, the aim of this study was to develop a new peak integration approach for the RP-HPLC patterns, in order to detect specific protein fingerprints and to allow a better differentiation between wheat cultivars within one species and between species. In that way, it was possible to achieve comprehensive insights into the changes of the protein distribution of wheat cultivars over the past 120 years due to breeding.

2 | MATERIALS AND METHODS

2.1 | Grain samples

Grains of 60 hexaploid German winter wheat cultivars from 1891 to 2010 were analyzed. For each decade, the five most common cultivars for the respective decade were selected (Table S1) and cultivated in a randomized field order without fertilization and harvested in 2015, 2016, and 2017 at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben, Germany (Pronin, Börner, Weber, & Scherf, 2020). Wholemeal flours were obtained by milling the grains with a laboratory grinder (Bosch, Stuttgart, Germany) and sieving to a particle size of 0.2 mm. Cultivars first registered prior to 1950 were designated as "old" (samples 1–30) and those registered after 1950 as "modern" (samples 31–60).

In addition, a second sample set containing different wheat species, that is, spelt, durum wheat, emmer, einkorn, and also common wheat, from a different region were analyzed (Table S1). Eight cultivars per wheat species were harvested in 2013 by the State Plant Breeding Institute (SPBI), University of Hohenheim (Stuttgart, Germany) at Seligenstadt, Germany. The cultivars were fertilized according to the demand of the particular wheat species. Common wheat, spelt, and durum wheat were fertilized with 95 kg N/ha and emmer and einkorn with 75 kg N/ha. In case of spelt,

emmer, and einkorn, the grains were dehulled. The grains were milled into wholemeal flour using a cross-beater mill (Perten Instruments, Hamburg, Germany) (Geisslitz, Wieser, Scherf, & Koehler, 2018; Longin et al., 2016).

2.2 | Characterization of the protein composition by RP-HPLC

The extraction of the wheat flours was performed stepwise based on solubility of the wheat protein fractions. First, the albumin/globulin fraction was obtained by extracting the wheat flours (100 mg) twice with 1.0 ml buffered salt solution (0.067 mol/L K_2HPO_4/KH_2PO_4-buffer, 0.4 mol/L NaCl, pH = 7.6). Next, the gliadin fraction was extracted three times with 0.5 ml 60% (v/v) aqueous ethanol from the residues. In the final step, the residues were extracted twice with 1.0 ml glutenin extraction solution (50% (v/v) 1-propanol, 0.1 mol/L Tris-HCl, pH 7.5, 0.06 mol/L (w/v) dithiothreitol) at 60°C under nitrogen, in order to obtain the glutenin fraction. The extraction conditions for each fraction were vortex mixing for 2 min at 22°C, followed by magnetic stirring for 10 min or 30 min (glutenins). After centrifugation of the suspensions (4,600 × *g*, 25 min, 22°C), the supernatants were combined, diluted to 2.0 ml with the respective extraction solvent, and filtered (Whatman™, Spartan 13/0.45 RC, GE Healthcare, Freiburg, Germany). The extractions for each sample were done in triplicate and analyzed by means of RP-HPLC according to Schalk et al., 2017.

For analysis, the following system was used: Jasco XLC system (Jasco, Gross-Umstadt, Germany) with ChromPass software (Jasco); column, Acclaim™ 300 C_{18} (particle size 3 µm, pore size 30 nm, 2.1 × 150 mm, Thermo Fisher Scientific, Braunschweig, Germany); temperature, 60°C; injection volume, 10 µl for gliadins and 20 µl for albumins/globulins and glutenins; elution solvents, trifluoroacetic acid (TFA) (0.1%, v/v) in water (A) and TFA (0.1%, v/v) in acetonitrile (B); linear gradient for albumins/globulins: 0 min 0% B, 0.5 min 20% B, 7 min 60% B, 7.1–11 min 90% B, 11.1–17 min 0% B; linear gradient for gliadins and glutenins: 0 min 0% B, 0.5 min 24% B, 20 min 56% B, 20.1–24.1 min 90% B, 24.2–30 min 0% B; flow rate 0.2 ml/min; and detection, UV absorbance at 210 nm. For external calibration, Prolamin Working Group (PWG)-gliadin (2.5 mg/ml in 60% ethanol) was used (van Eckert et al., 2006).

2.3 | Integration methods

Peak area integration for each protein fraction was performed by the ChromPass software. Following the traditional approach, the albumin/globulin fraction is integrated as one, starting at the retention time of 5.6 min and terminating at 12.9 min. The integration of the gliadin fraction is divided into ω5-gliadins at 7.8–10.0 min, ω1,2-gliadins at 10.0–13.4 min, α-gliadins at 13.4–18.0 min, and γ-gliadins at 18.0–21.7 min. Concerning the glutenin fraction, the integration starts with ωb-gliadins at 7.5–10.0 min, followed by HMW-GS at 10.0–14.5 min and LMW-GS at 14.5–22.7 min.

In case of the new integration approach, integration events were carried out every 20 s for all protein fractions. For albumins/globulins, 22 integration events were performed, from 5.6 min to 12.9 min. For gliadins, 42 integration events were performed from 7.8 min to 21.7 min and the integration for glutenins started at 7.5 min and ended at 22.7 min, which makes 46 integration events in total. For the determination of the area proportions (%) for each integration event, the ratio of the area beneath the peak relative to the area of the total chromatogram was calculated.

2.4 | Statistical analysis

Principal component analysis (PCA) was used to differentiate between old and modern wheat cultivars. For the traditional integration strategy, the percentages of albumins/globulins, gliadins (ω5-, ω1,2-, α-, γ-gliadins) and glutenins (ωb-gliadins, HMW-GS, LMW-GS) in the total RP-HPLC protein content for each of the 60 wheat cultivars were correlated with each other per harvest year. Additionally, the percentages were averaged over three harvest years and also correlated with each other. Using the new integration approach, the percentages of all integration events in the total RP-HPLC protein content for each of the 60 wheat cultivars were correlated with each other, both per each harvest year and averaged over three harvest years. PCA was carried out using Excel with the XLSTAT statistical software (Addinsoft, New York, NY, USA). Hierarchical cluster analysis was performed using the Origin 2018b software (OriginLab, Northampton, Massachusetts, USA), and the corresponding data were normalized (0;1) before distance calculation for clustering observations. The correlation coefficients (r) were defined as $r \leq \pm.54$ no correlation, $\pm.54 < r \leq \pm.67$ weak correlation, $\pm.67 < r \leq \pm.78$ medium correlation, and $r > \pm.78$ strong correlation (Thanhaeuser, Wieser, & Koehler, 2014).

3 | RESULTS AND DISCUSSION

3.1 | Distinction of wheat cultivars using the traditional integration method

The traditional way of peak integration for cereal proteins as described in Wieser et al. (1998) is as follows: The albumin/globulin fraction is not subdivided further, while the gliadin and glutenin fractions are subdivided into ω5-, ω1,2-, α-, and

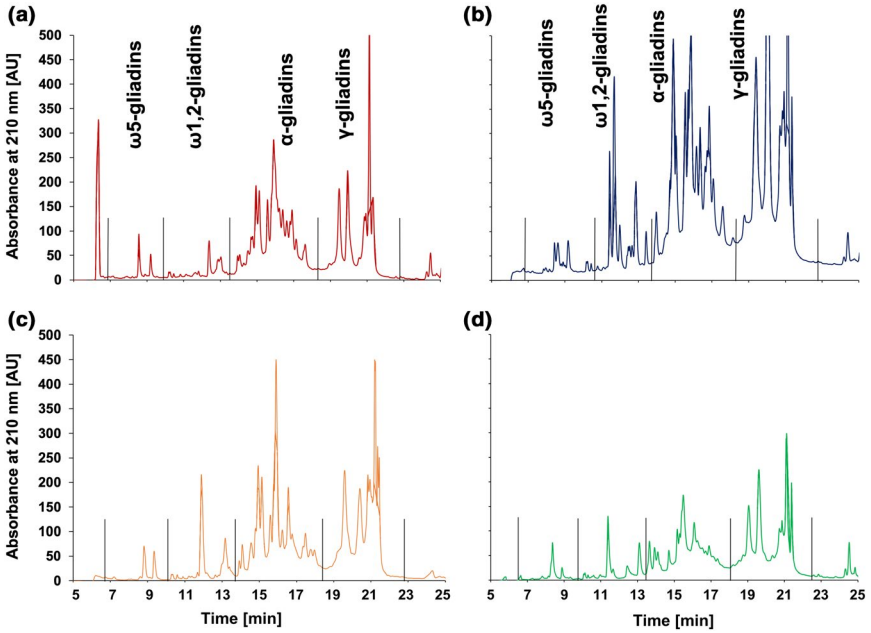

FIGURE 1 Reversed-phase (RP)-HPLC chromatograms of the gliadin fractions of the samples 1 (a), 24 (b), 35 (c) 58 (d) from the harvest year 2015 [Colour figure can be viewed at wileyonlinelibrary.com]

γ-gliadins as well as ωb-gliadins, HMW-GS, and LMW-GS, respectively (Figure 1). A PCA was performed in order to reduce the dimensionality of the data set and to see possible changes between old and modern wheat cultivars. The percentages of albumins/globulins, ω5-, ωb-, ω1,2-, α-, and γ-gliadins, HMW-GS, and LMW-GS relative to the sum of extractable proteins for each of the 60 common wheat samples of each harvest year (Figure S1) and of the average over three years were correlated with each other (Figure 2a). The principal components (PC)1 and PC2 accounted for 57.84% and 17.61% of the observed variability, respectively. PC1 was positively correlated with γ-gliadins ($r = .775$) and with gliadins and gliadin/glutenin ratios ($r \geq .978$), but negatively correlated with HMW-GS ($r = -.849$), glutenins, and LMW-GS ($r \geq -.971$). In contrast, PC2 was only negatively correlated with albumins/globulins ($r = -.868$). A negative correlation ($r = -.885$) was observed between gliadins and glutenins, but there was no correlation between ω5-, ω1,2-gliadins and gliadins or glutenins. Further, there was no correlation between albumins/globulins and gliadins or glutenins. The majority of the old wheat cultivars was placed on the right, because of their high gliadin and comparatively low glutenin proportions, as also reflected in the gliadin/glutenin ratios. Vice versa, the majority of the modern wheat cultivars was placed on the left, because of their high glutenin and low gliadin proportions. Neither old nor modern wheat cultivars were related to albumin/globulin proportions. Both old and modern cultivars showed a cluster formation, but with clear overlap in the middle of the biplot and several exceptions. The PCA revealed that the old cultivars 20, 22, and 26 were located among modern cultivars because of their exceptionally high glutenin proportions. Moreover, the modern wheat cultivars 31 and 44 were always located among old cultivars, because of their relatively high gliadin proportions. Essentially, the PCA of each of the three harvest years showed comparable correlations regarding wheat cultivars and corresponding proportions of the protein fractions. A further difference between old and modern cultivars was that the respective old cultivars of one decade were distributed unevenly, whereas the majority of the modern wheat cultivars showed cluster formation. Thus, it appears that the distinction between old and modern cultivars is limited using the traditional integration method, because information on individual peaks is lost.

3.2 | Qualitative characterization of the protein composition by RP-HPLC

By comparing all chromatograms of the 60 samples, significant changes in the peak patterns were determined over the period of 120 years. We identified cultivars with distinct patterns both for old and modern cultivars, which were 1, 24 and 35 and 58. Figure 1 shows RP-HPLC profiles of the corresponding gliadin fractions (harvest year 2015, overlay in Figure S2) and these agreed well with previous reports (Eggert, Wieser, & Pawelzik, 2010; Geisslitz et al., 2018; Schalk et al., 2017).

Due to comparatively clear baseline separation and presence of few peaks, the most noticeable differences were observed for the ω5- and ω1,2-gliadins. Compared to sample 1 (Figure 1a), the RP-HPLC profile of sample 24 (Figure 1b) revealed one additional peak for the ω5-gliadins and two additional high intensity peaks for ω1,2-gliadins at retention times of 8.5–9.0 min and 11.0–12.5 min, respectively. Sample 35 (Figure 1c) contained two sharp and baseline separated peaks for the ω5-gliadins at the retention times 8.5–10

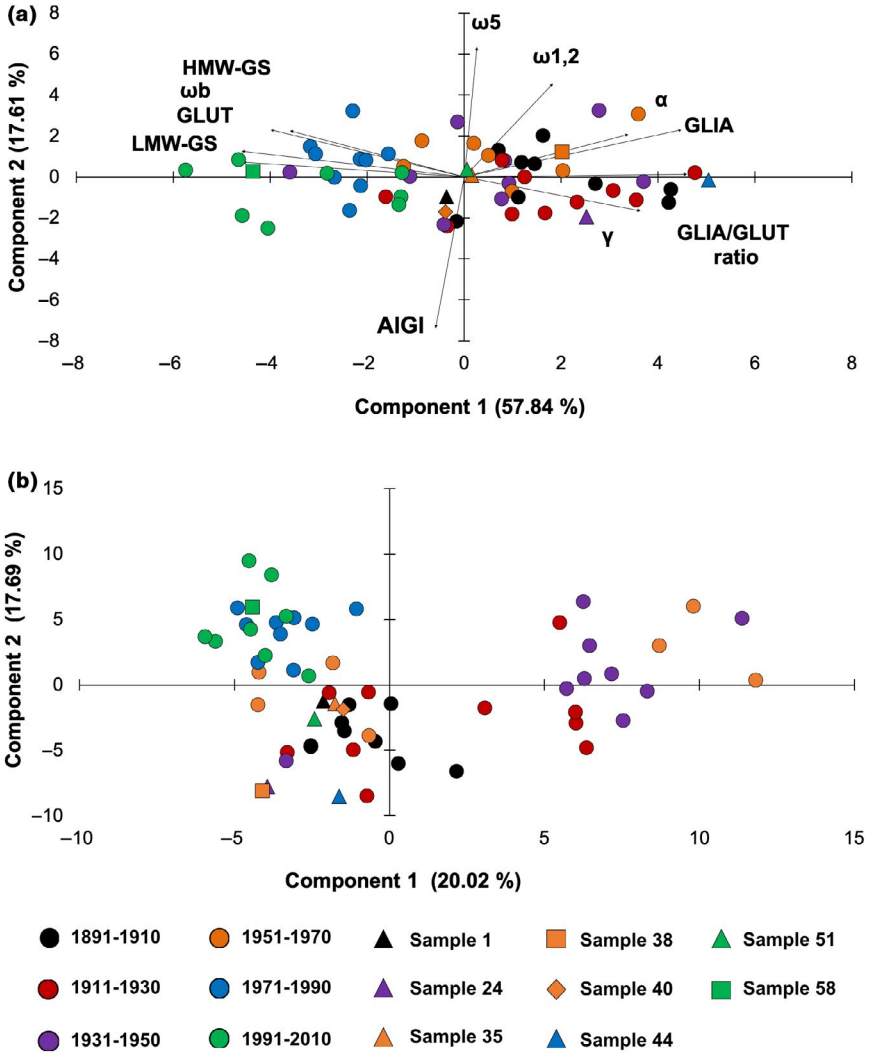

FIGURE 2 PCA biplot of proportions of albumins/globulins, ω5-, ω1,2-, α-, γ-gliadins, ωb-gliadins, HMW-GS, and LMW-GS relative to the sum of extractable proteins (a) and PCA biplot of proportions of each integration event of albumins/globulins, gliadins, and glutenins relative to the sum of extractable proteins (b). The data are displayed as average of the three harvest years 2015, 2016, and 2017. Distinct samples 1, 24, 35, 38, 40, 44, 51, and 58 were marked with different symbols. PCA, principal component analysis [Colour figure can be viewed at wileyonlinelibrary.com]

min and one sharp and one shoulder peak for ω1,2-gliadins at 11.0–13.8 min, respectively. In contrast, sample 58 (Figure 1d) showed an additional peak for the ω1,2-gliadins between 11.0 and 13.8 min. Further differences in the peak distribution were observed for the patterns of α-gliadins. Especially for the samples 24 and 35, it was clearly visible that certain peaks were more pronounced than comparable ones at the same retention time. The patterns of γ-gliadins mostly remained unchanged, except for one peak at the retention time of 20 min, which changed its intensity from sample to sample. However, the detailed characterization of the patterns of α- and γ-gliadins was limited due to their accumulation and complexity of peaks and the absence of a baseline separation. Nonetheless, it was striking that the RP-HPLC profiles have significantly changed from sample 1 to 60 in terms of peak variability and distribution. Our findings did not support the results of Malalgoda, Ohm, Meinhardt, and Simsek (2017), who reported no changes in the profiles of hard red spring wheats between 1910 and 2013. In fact, we found that there seem to be differences between the profiles of old and modern cultivars. Interestingly, the profile changes were not reflected in the contents of the protein fractions and types (Pronin et al., 2020). This is why a new approach of peak integration was required to allow a better distinction of different wheat cultivars and an advanced detection of the protein fingerprints.

3.3 | Distinction of wheat cultivars using the new integration method

The new method involves continuous integration every 20 s, regardless of the protein fraction. In this way, nearly every peak was recognized individually (Figure S3). Compared to the traditional classification, peaks 1–7 belong to ω5-gliadins, 8–18 to ω1,2-gliadins, 19–33 to α-gliadins, and 34–42 to γ-gliadins. Concerning the glutenin fraction, peaks 1–9 belong to ωb-gliadins, 10–22 to HMW-GS, and 23–46 to LMW-GS.

Principal component analysis biplots were created by correlating all proportions of each integration event of albumins/globulins, gliadins, and glutenins relative to the sum of

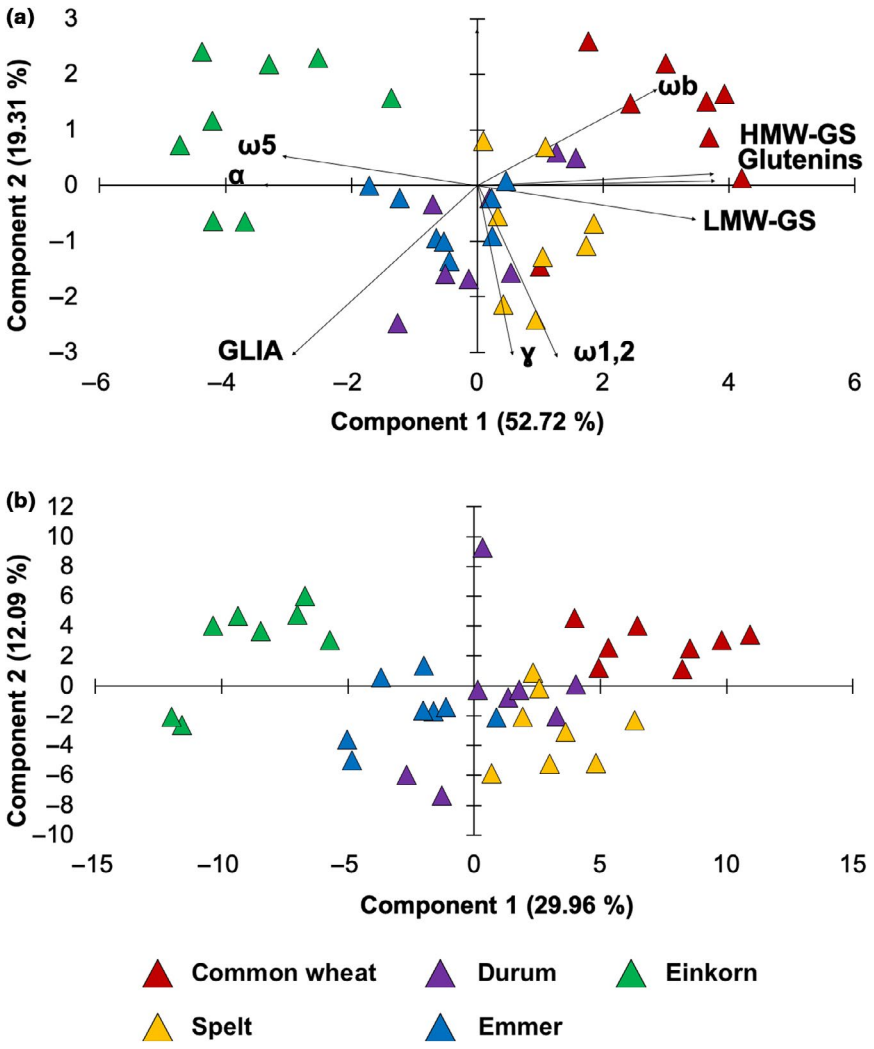

FIGURE 3 PCA biplot of proportions of albumins/globulins, ω5-, ω1,2-, α-, γ-gliadins, ωb-gliadins, HMW-GS, and LMW-GS relative to the sum of extractable proteins (a) and PCA biplot of proportions of each integration event of albumins/globulins, gliadins, and glutenins in the sum of extractable proteins from common wheat, spelt, durum, emmer, and einkorn (b). PCA, principal component analysis [Colour figure can be viewed at wileyonlinelibrary.com]

extractable proteins. At first glance, it becomes obvious that the harvest year had a strong influence on the cluster formation and distribution of the cultivars (Figure S4), which is in accordance with our previous study (Pronin et al., 2020) and also evident using the traditional integration method. One reason for the stronger clustering of the wheat cultivars might be that this year had the least amount of annual mean precipitation (366 mm, as opposed to 533 mm and 557 mm in 2015 and 2017, respectively), but further investigations would be necessary to establish a clear cause–effect relation here. However, to even out the environmental effects, the results of the average over three harvest years are discussed in the following (Figure 2b). PC1 and PC2 accounted for 20.02% and 17.69% of the observed variability, respectively. PC1 was positively correlated with gliadin peaks 17–19 ($r \geq .81$) and with glutenin peaks 9 and 10 ($r = .83$ and .89). On the contrary, PC2 was only positively correlated with glutenin peaks 4 and 20–22 ($r \geq .72$), but negatively with total gliadin proportions ($r = -.76$) and gliadin peaks 24, 30–41 ($r \geq -.72$). Furthermore, we observed a cluster formation for modern cultivars (cluster I), which were located in the upper left corner, whereas old cultivars were grouped into two clusters. One cluster was placed on the right side and mainly contained samples, which encompassed the transition from old to modern cultivars including modern samples 31, 32, and 34 (cluster II). The other cluster also included both old and a few modern samples (35, 38, 40, 44, 51) and was placed in the bottom left corner (cluster III). Moreover, we observed a cluster formation for five cultivars of each decade, except the cultivars listed above. Considering the correlation between the three clusters and protein peaks, we observed clear determinants for each cluster, which were also true over all three harvest years. First, modern cultivars (cluster I) were correlated with peaks for albumins/globulins (2–5 and 16–22), gliadins (3, 35, and 37), and the majority of glutenins (4–7, 15–20, and 32–43). Next, old cultivars (cluster II) were correlated with peaks for albumins/globulins (7, 8, 10, 11), gliadins (1, 2, 5–10 15–23), and glutenins (2, 3, 8–11 13, 14, 23–27, 29, 30). This cluster was barely correlated with the fractions, in particular only with peaks of albumins/globulins (12–15), gliadins (4, 13, 14, 25–29, and 36–42), and only with a few of glutenins (37, 44, 46). Thus, old cultivars were

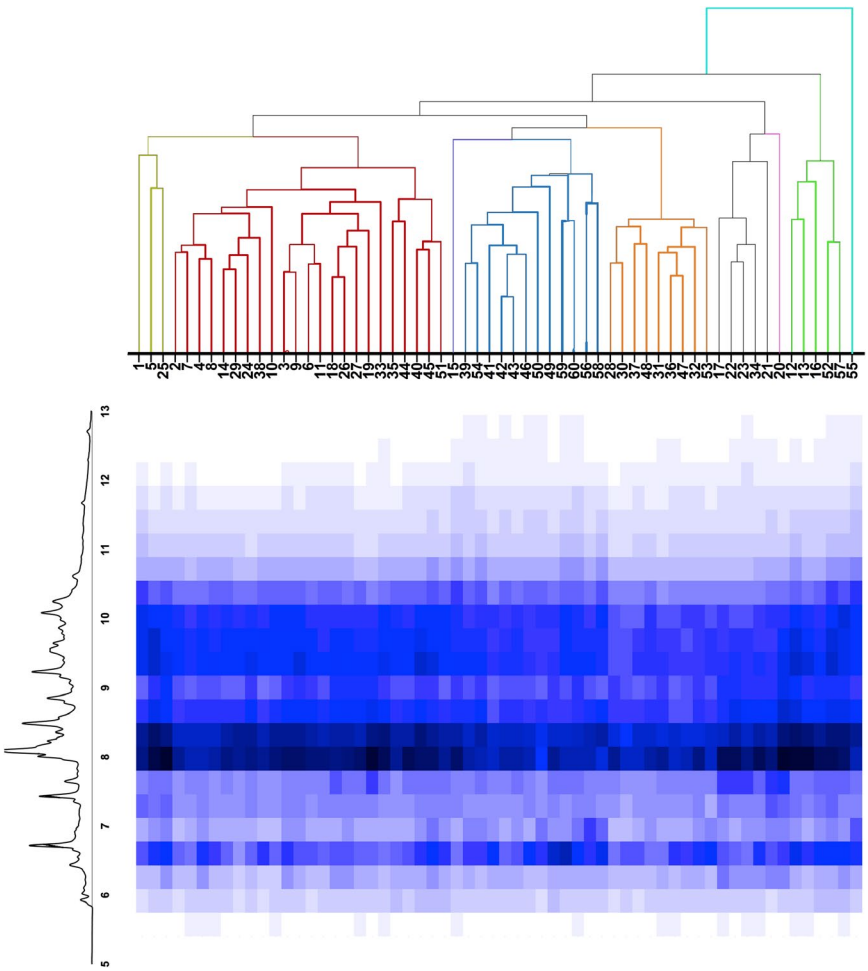

FIGURE 4 Dendrogram and heatmap based on RP-HPLC data for albumins/globulins using the new integration approach. The dendrogram on top shows the 60 analyzed wheat cultivars, and the qualitative profile on the side shows the pattern of the corresponding protein fraction. The data are displayed as average of the three harvest years 2015, 2016, and 2017. RP, reversed-phase [Colour figure can be viewed at wileyonlinelibrary.com]

mostly associated with the middle part of albumins/globulins, the front parts of ω5- and α-gliadins, and the back part of ω1,2-gliadins. Further, old cultivars were associated with ωb-gliadins, the front and middle parts of HMW-GS, and the front part of LMW-GS. On the other hand, modern cultivars were mostly related to glutenins, especially to the back parts of ω5- and ωb-gliadins, HMW-GS and LMW-GS, to the front part of albumins/globulins and the front part of γ-gliadins. Cluster III was mostly influenced by the back parts of albumins/globulins, ω5-gliadins and α-gliadins and by the total of ω1,2- and γ-gliadins. The observations made for the average over three harvest years were influenced by the harvest year 2016 (Figure S4b), since the clusters for the harvest years 2015 and 2017 were mostly overlapping and therefore difficult to distinguish. These observations are consistent with the fact the harvest year had a significant effect on the protein distribution of the 60 wheat cultivars (Pronin et al., 2020). Effectively, there is a lack of studies performing a comprehensive comparison of wheat protein patterns by means of RP-HPLC, except for the study by Daniel and Triboi (2000) that revealed significant differences between the gliadin chromatograms of a winter wheat cultivar influenced by the effects of temperature and fertilization.

3.4 | Distinction of wheat species using the new integration method

To test the applicability of the new integration method, a second sample set including the five wheat species common wheat, spelt, durum wheat, emmer, and einkorn was analyzed as described above (Table S1). We observed a distinct cluster formation for the cultivars of common wheat and einkorn, but clear overlap between spelt, durum wheat, and emmer with both integration methods (Figure 3). The new integration performed better in terms of distinguishing emmer from spelt, but the durum wheat samples still lay in between, because they showed the greatest within-species variability. As expected, common wheat was associated with glutenins (Geisslitz et al., 2018), whereas ω5- and α-gliadins mostly determined the placement of einkorn cultivars. The results are in agreement with the PCA reported by Geisslitz et al. (2019) that also showed an association between glutenin contents and the location of common wheat cultivars, as well as between gliadin contents and einkorn. Our new approach allowed a better differentiation of wheat species than before, but there were also similarities regarding the overlap of spelt, durum wheat, and emmer.

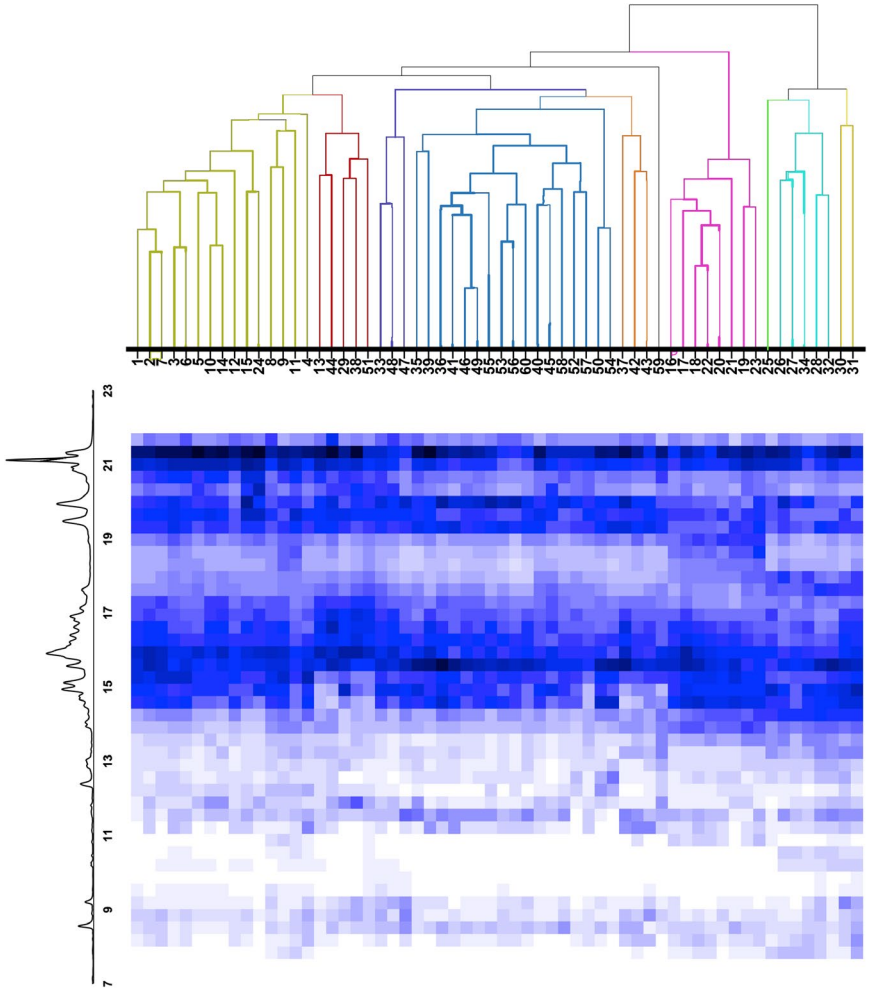

FIGURE 5 Dendrogram and heatmap based on RP-HPLC data for gliadins using the new integration approach. The dendrogram on top shows the 60 analyzed wheat cultivars, and the qualitative profile on the side shows the pattern of the corresponding protein fraction. The data are displayed as average of the three harvest years 2015, 2016, and 2017. RP, reversed-phase [Colour figure can be viewed at wileyonlinelibrary.com]

3.5 | Hierarchical cluster analysis

To determine the differences and similarities for the cultivars in more detail, the data for each protein fraction were clustered by similarity using hierarchical clustering analysis (Figures 4–6). The following characterization of the cultivars refers to the average of the three harvest years. First, the dendrogram of the albumins/globulins shows that the 60 cultivars can be divided into nine major groups according to their distribution (Figure 4). The most modern samples (56–60) showed a strong correlation with fairly modern samples (41–49) and were also correlated with samples of the transition from old to modern (28, 30, 31). Further, the samples from each decade were often distributed unevenly over the 9 groups, but modern samples showed more similarities than old cultivars. We further observed individual cultivars, that did not appear jointly with the cultivars from the same decade, but were found in different groups and showed greater distances to the cultivars from the same decade, for example, samples 18, 19, 24, 25, 29, and 38. Interestingly, these cultivars showed rather short distances among each other. In detail, we found samples which appeared to be different within one group, because they belonged to other decades, for example, the second group contained modern samples 33, 35, 40, 44, 45, and 51, which were similar to each other based on their distance, but were located among old cultivars within the group. Interestingly, these cultivars were always correlated irrespective of the harvest year and were not among the other cultivars from the same decade. What is more, the modern sample 38 was found in group 2, among old cultivars and with no correlation to the samples from the corresponding decade. Further outstanding samples, which showed disparities within the group on the one hand and between the group on the other, were 15 (group 3); 47, 48, 53 (group 5); 34 (group 6); 20 (group 7); 52, 57 (group 8) and 55 (group 9).

Next, the cultivars of the gliadin fraction were categorized into 10 groups (Figure 5). It was noticeable that old cultivars of the particular decade from sample 1 to 25, except for sample 24, were mostly located in the same group, while cultivars from the 1940s and 1950s were rather evenly distributed. Especially, the samples 29, 33, and 35 showed greater distances to the cultivars from the same period. Modern samples were widely distributed, but remained in the same or in an adjoining section (groups 3 and 4). Exceptions with wider distances to the modern cultivars from the same decade were observed for samples 38, 44, 51,

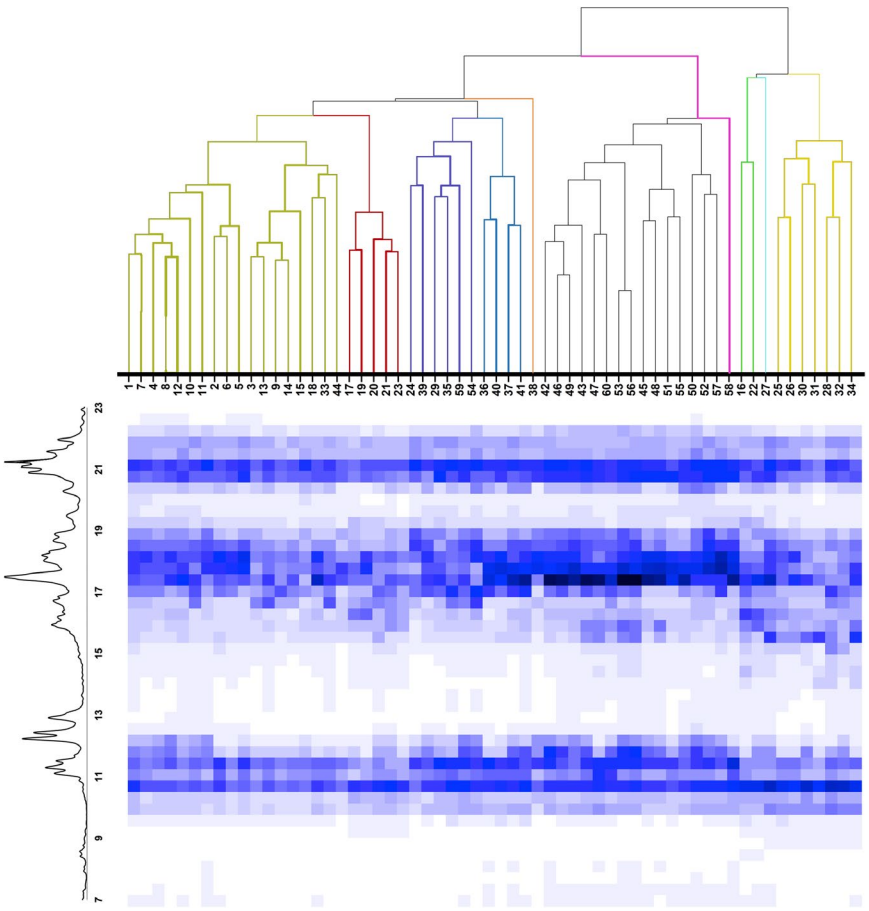

FIGURE 6 Dendrogram and heatmap based on RP-HPLC data for glutenins using the new integration approach. The dendrogram on top shows the 60 analyzed wheat cultivars, and the qualitative profile on the side shows the pattern of the corresponding protein fraction. The data are displayed as average of the three harvest years 2015, 2016, and 2017. RP, reversed-phase [Colour figure can be viewed at wileyonlinelibrary.com]

and 59. However, we found that the oldest samples (1–15) were most likely correlated with modern samples (36–60) and not with cultivars from the 1940s and 1950s (26–35). Noticeable samples were also found for the gliadin fraction, in particular 24 (group 1); 29, 38, 44, 51 (group 2); 33 (group 3); 59 (group 6) and 25 (group 8).

Last, the cultivars of the glutenin fraction were categorized into 10 groups (Figure 6). Old and modern samples, respectively, were mostly correlated among themselves. It was striking that the samples of the same decade were found in the same or adjacent group, but with a few exceptions. These exceptions were 33, 44 (group 1); 59 (group 3). Effectively, samples 38 (group 5); 58 (group 7) and 27 (group 8) showed great distances and were not correlated to other groups.

The genealogical relations between the cultivars were obtained from the Genetic Resources Information System (GRIS) database, but some cultivars were not included (e.g., 9, 12, 23, 30) and some are directly derived from German landraces (e.g., 8, 10, 14 and 20). Some cultivars included in the sample set did have relations, for example, 59 has 31 and 43 in its close ancestry, but these were only partially reflected. While 43 was next to 59 in the gliadin dendrogram, 31 was not; and in the glutenin dendrogram these three were not any closer to one another than to totally unrelated cultivars. It seems that the resolution of the RP-HPLC chromatograms is not sufficiently detailed enough to map genealogical relations or individual protein isoforms.

Overall, it can be concluded that the observations above revealed associations among cultivars and emphasized the uniqueness of particular cultivars. We were able to find one cultivar, which was outstanding with respect to its unique albumin/globulin and gliadin pattern and moreover three further cultivars which had unique patterns for all protein fractions. These samples were 59 and 33, 38, and 44, respectively. In a recent study, 30 hard red spring wheat cultivars from a period of 100 years were analyzed to elucidate the clustering patterns of gliadins by means of RP-HPLC (Malalgoda et al., 2017). The data were produced by determining the presence and absence time of a particular peak for each cultivar and indicated that the cultivars have in fact changed over the last 100 years in terms of quality characteristics. However, no clusters according to the release year of the cultivars were observed. In the view of the above, the insights given by the new method of peak integration are crucial for an in-depth analysis of wheat proteins. The major advantage of this method is that the distribution of individual peaks in the complex gluten chromatograms is now captured more comprehensively, which leads to better results. For future research, these insights can be used to isolate specific

peaks from the outstanding cultivars for in-depth MS and immunological studies.

4 | CONCLUSION

The new integration approach for wheat protein fingerprints analyzed by means of RP-HPLC was suitable to trace changes in peak variability depending on the wheat cultivar or wheat species. The method was also applicable to spelt, emmer, durum wheat, and einkorn compared to common wheat, although the traditional method of peak integration performed almost equally well in this case. Within the set of 60 common wheat cultivars, we were able to differentiate between old and modern wheat cultivars and additionally identify unique cultivars that stood out because of their protein composition. Four samples were identified that warrant further research to identify the specific proteins that are responsible for the differences.

ACKNOWLEDGEMENTS

The authors thank Alexandra Axthelm, Angelika Grassl, and Stefanie Hackl (Leibniz-LSB@TUM) for excellent technical assistance. Special thanks go to C. Friedrich H. Longin from the State Plant Breeding Institute, University of Hohenheim, Germany, for cultivating and providing the set of common wheat, spelt, durum wheat, emmer, and einkorn samples also used in this study. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no competing financial interest.

ORCID

Sabrina Geisslitz https://orcid.org/0000-0002-6811-4069
Andreas Börner https://orcid.org/0000-0003-3301-9026
Katharina A. Scherf https://orcid.org/0000-0001-8315-5400

REFERENCES

Anjum, F. M., Khan, M. R., Din, A., Saeed, M., Pasha, I., & Arshad, M. U. (2007). Wheat gluten: High molecular weight glutenin subunits – Structure, genetics, and relation to dough elasticity. *Journal of Food Science*, *72*, R56–R63. https://doi.org/10.1111/j.1750-3841.2007.00292.x

Bin, Q., & Peterson, D. G. (2016). Identification of bitter compounds in whole wheat bread crumb. *Food Chemistry*, *203*, 8–15. https://doi.org/10.1016/j.foodchem.2016.01.116

Bromilow, S. N., Gethings, L. A., Langridge, J. I., Shewry, P. R., Buckley, M., Bromley, M. J., & Mills, E. N. (2017). Comprehensive proteomic profiling of wheat gluten using a combination of data-independent and data-dependent acquisition. *Frontiers in Plant Science*, *7*, 2020. https://doi.org/10.3389/fpls.2016.02020

Colgrave, M. L., Goswami, H., Byrne, K., Blundell, M., Howitt, C. A., & Tanner, G. J. (2015). Proteomic profiling of 16 cereal grains and the application of targeted proteomics to detect wheat contamination. *Journal of Proteome Research*, *14*, 2659–2668. https://doi.org/10.1021/acs.jproteome.5b00187

Daniel, C., & Triboi, E. (2000). Effects of temperature and nitrogen nutrition on the grain composition of winter wheat: Effects on gliadin content and composition. *Journal of Cereal Science*, *32*, 45–56. https://doi.org/10.1006/jcrs.2000.0313

Davis, K. R., Cain, R. F., Peters, L. J., Le Tourneau, D., & McGinnis, D. (1981). Evaluation of the nutrient composition of wheat. II. Proximate analysis, thiamin, riboflavin, niacin, and pyridoxine. *Cereal Chemistry*, *58*, 116–120.

Delcour, J. A., Joye, I. J., Pareyt, B., Wilderjans, E., Brijs, K., & Lagrain, B. (2012). Wheat gluten functionality as a quality determinant in cereal-based food products. *Annual Review of Food Science and Technology*, *3*, 469–492. https://doi.org/10.1146/annurev-food-022811-101303

Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat. *Science*, *316*, 1862–1866. https://doi.org/10.1126/science.1143986

DuPont, F. M., Vensel, W., Chan, R., & Kasarda, D. D. (2000). Characterization of the 1B-type ω-gliadins from *Triticum aestivum* cultivar Butte. *Cereal Chemistry*, *77*, 607–614. https://doi.org/10.1094/CCHEM.2000.77.5.607

DuPont, F. M., Vensel, W., Encarnacao, T., Chan, R., & Kasarda, D. D. (2004). Similarities of omega gliadins from *Triticum urartu* to those encoded on chromosome 1A of hexaploid wheat and evidence for their post-translational processing. *Theoretical and Applied Genetics*, *108*, 1299–1308. https://doi.org/10.1007/s00122-003-1565-9

Dupont, F. M., Vensel, W. H., Tanaka, C. K., Hurkman, W. J., & Altenbach, S. B. (2011). Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry. *Proteome Science*, *9*, 10. https://doi.org/10.1186/1477-5956-9-10

Eggert, K., Wieser, H., & Pawelzik, E. (2010). The influence of *Fusarium* infection and growing location on the quantitative protein composition of (part I) emmer (*Triticum dicoccum*). *European Food Research and Technology*, *230*, 837–847. https://doi.org/10.1007/s00217-010-1229-3

Geisslitz, S., America, A. H., & Scherf, K. A. (2020). Mass spectrometry of in-gel digests reveals differences in amino acid sequences of high-molecular-weight glutenin subunits in spelt and emmer compared to common wheat. *Analytical and Bioanalytical Chemistry*, *412*, 1277–1289. https://doi.org/10.1007/s00216-019-02341-9

Geisslitz, S., Longin, C. F. H., Scherf, K. A., & Koehler, P. (2019). Comparative study on gluten protein composition of ancient (einkorn, emmer and spelt) and modern wheat species (durum and common wheat). *Foods*, *8*, 409. https://doi.org/10.3390/foods8090409

Geisslitz, S., Wieser, H., Scherf, K., & Koehler, P. (2018). Gluten protein composition and aggregation properties as predictors for bread volume of common wheat, spelt, durum wheat, emmer and einkorn. *Journal of Cereal Science*, *83*, 204–212. https://doi.org/10.1016/j.jcs.2018.08.012

Kasarda, D. D. (2013). Can an increase in celiac disease be attributed to an increase in the gluten content of wheat as a consequence of wheat breeding? *Journal of Agricultural and Food Chemistry*, *61*, 1155–1159. https://doi.org/10.1021/jf305122s

Lagrain, B., Rombouts, I., Wieser, H., Delcour, J. A., & Koehler, P. (2012). A reassessment of the electrophoretic mobility of high

molecular weight glutenin subunits of wheat. *Journal of Cereal Science*, *56*, 726–732. https://doi.org/10.1016/j.jcs.2012.08.003

Longin, C. F. H., Ziegler, J., Schweiggert, R., Koehler, P., Carle, R., & Würschum, T. (2016). Comparative study of hulled (einkorn, emmer, and spelt) and naked wheats (durum and bread Wheat): Agronomic performance and quality traits. *Crop Science*, *56*, 302. https://doi.org/10.2135/cropsci2015.04.0242

Malalgoda, M., Ohm, J.-B., Meinhardt, S., Chao, S., & Simsek, S. (2017). Cluster analysis of historical and modern hard red spring wheat cultivars based on parentage and HPLC analysis of gluten-forming proteins. *Cereal Chemistry*, *94*, 560–567. https://doi.org/10.1094/CCHEM-08-16-0223-R

Pronin, D., Börner, A., Weber, H., & Scherf, K. A. (2020). Wheat (*Triticum aestivum* L.) breeding from 1891 to 2010 contributed to increasing yield and glutenin contents but decreasing protein and gliadin contents. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.0c02815

Schalk, K., Lexhaller, B., Koehler, P., & Scherf, K. A. (2017). Isolation and characterization of gluten protein types from wheat, rye, barley and oats for use as reference materials. *PLoS One*, *12*, e0172819. https://doi.org/10.1371/journal.pone.0172819

Scherf, K. A., Koehler, P., & Wieser, H. (2016). Gluten and wheat sensitivities – An overview. *Journal of Cereal Science*, *67*, 2–11. https://doi.org/10.1016/j.jcs.2015.07.008

Shewry, P. R., Miflin, B. J., & Kasarda, D. D. (1984). The structural and evolutionary relationships of the prolamin storage proteins of barley, rye and wheat. *Philosophical Transactions of the Royal Society B*, *304*, 297–308. https://doi.org/10.1098/rstb.1984.0025

Shewry, P. R., Piironen, V., Lampi, A. M., Edelmann, M., Kariluoto, S., Nurmi, T., … Ward, J. L. (2010). The HEALTHGRAIN wheat diversity screen: Effects of genotype and environment on phytochemicals and dietary fiber components. *Journal of Agricultural and Food Chemistry*, *58*, 9291–9298. https://doi.org/10.1021/jf100039b

Thanhaeuser, S. M., Wieser, H., & Koehler, P. (2014). Correlation of quality parameters with the baking performance of wheat flours. *Cereal Chemistry*, *91*, 333–341. https://doi.org/10.1094/CCHEM-09-13-0194-CESI

van Eckert, R., Berghofer, E., Ciclitira, P. J., Chirdo, F., Denery-Papini, S., Ellis, H. J., … Wieser, H. (2006). Towards a new gliadin reference material-isolation and characterisation. *Journal of Cereal Science*, *43*, 331–341. https://doi.org/10.1016/j.jcs.2005.12.009

Ward, J. L., Poutanen, K., Gebruers, K., Piironen, V., Lampi, A., Nyström, L., … Shewry, P. (2008). The HEALTHGRAIN Cereal Diversity Screen: Concept, results, and prospects. *Journal of Agricultural and Food Chemistry*, *12*, 9699–9709. https://doi.org/10.1021/jf8009574

Wieser, H., Antes, S., & Seilmeier, W. (1998). Quantitative determination of gluten protein types in wheat flour by reversed-phase high-performance liquid chromatography. *Cereal Chemistry*, *75*, 644–650. https://doi.org/10.1094/CCHEM.1998.75.5.644

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pronin D, Geisslitz S, Börner A, Scherf KA. Fingerprinting of wheat protein profiles for improved distinction between wheat cultivars and species. *Cereal Chem*. 2020;97:999–1009. https://doi.org/10.1002/cche.10323