

## ARTICLE OPEN



# Reference proteomes of five wheat species as starting point for future design of cultivars with lower allergenic potential

Muhammad Afzal<sup>1,5</sup>, Malte Sielaff<sup>2,5</sup>, Ute Distler<sup>2</sup>, Detlef Schuppan<sup>3,4,6</sup>, Stefan Tenzer<sup>1,6</sup> and C. Friedrich H. Longin<sup>1,6</sup>✉

Wheat is an important staple food and its processing quality is largely driven by proteins. However, there is a sizable number of people with inflammatory reactions to wheat proteins, namely celiac disease, wheat allergy and the syndrome of non-celiac wheat sensitivity. Thus, proteome profiles should be of high importance for stakeholders along the wheat supply chain. We applied liquid chromatography-tandem mass spectrometry-based proteomics to establish the flour reference proteome for five wheat species, ancient to modern, each based on 10 cultivars grown in three diverse environments. We identified at least 2540 proteins in each species and a cluster analyses clearly separated the species based on their proteome profiles. Even more, >50% of proteins significantly differed between species - many of them implicated in products' quality, grain-starch synthesis, plant stress regulation and proven or potential allergic reactions in humans. Notably, the expression of several important wheat proteins was found to be mainly driven by genetics vs. environmental factors, which enables selection and refinement of improved cultivars for the wheat supply chain as long as rapid test methods will be developed. Especially einkorn expressed 5.4 and 7.2-fold lower quantities of potential allergens and immunogenic amylase trypsin inhibitors, respectively, than common wheat, whereas potential allergen content was intermediate in tetraploid wheat species. This urgently warrants well-targeted clinical studies, where the developed reference proteomes will help to design representative test diets.

npj Science of Food (2023)7:9; <https://doi.org/10.1038/s41538-023-00188-0>

## INTRODUCTION

Wheat is one of the most important staple foods with a worldwide production of 765 million tons in 2019 (<https://www.fao.org/faostat/en/#data/QCL>, accessed on 08.12.2021) and provides 20% of the daily intake of dietary protein together with fiber, minerals and vitamins<sup>1</sup>. Most of the production is contributed by modern species common wheat (*Triticum aestivum* ssp. *aestivum*) and durum (*Triticum turgidum* ssp. *durum*). While common wheat is cultivated globally on almost 223 million hectares (<https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery>, accessed on 29.01.2022) for bread production and animal nutrition, durum wheat covers worldwide 16 million hectares primarily for pasta production<sup>2</sup>. Although, ancient species spelt (*Triticum aestivum* ssp. *spelta*), emmer (*Triticum turgidum* ssp. *dicoccum*) and einkorn (*Triticum monococcum* ssp. *monococcum*) have been utilized as food for thousands of years<sup>3–5</sup>, they are currently cultivated only on a small-scale confined to specific regions<sup>6–8</sup>. Wheat grains contain roughly 8–15% protein of dry weight<sup>9</sup>, which can be classified into albumins/globulins (15–20%), including important essential amino acids<sup>10</sup>, and gluten proteins (80–85%)<sup>9,11</sup>. The viscoelastic and gustatory attributes as hallmarks of the quality for bread and pasta production are mainly endowed by the gluten proteins<sup>12,13</sup>. Simultaneously, some wheat proteins can trigger inflammatory reactions such as celiac disease (CeD), classical wheat allergy (WA), and non-celiac wheat sensitivity (NCWS) in approximately 1%, below 1%, and up to 10% of the wheat-consuming populations, respectively<sup>14</sup>. Specific gluten peptide sequences cause CeD<sup>15,16</sup>, alpha-amylase/trypsin inhibitors (ATIs) stimulate innate immune cells via toll-like receptor

4 (TLR4) to promote intestinal and extraintestinal inflammation in animal models of disease<sup>17–23</sup>, and especially serpins, lipid transfer proteins (LTPs),  $\beta$ -amylases, ATIs and some gluten proteins can cause immediate-type immunoglobulin E (IgE) mediated allergic reactions<sup>24–29</sup>. Moreover, clinical and functional studies suggest that type 2 food allergies, driven by, e.g., eosinophils and prominently to wheat proteins, play an important role in promoting irritable bowel syndrome<sup>14,30,31</sup>.

Previous studies compared modern and ancient wheat species limited to proteins of a specific family such as gluten proteins<sup>32</sup> or ATIs<sup>33–35</sup>, or studied the immunogenic potential of ATIs between common wheat and einkorn<sup>21,36</sup>. Compared to earlier gel-based proteomic studies, latest developments in LC-MS-based proteomics allow to quantify thousands of proteins in less than 1.5–2 h per sample<sup>37</sup>. Recently, this proteomic technology was applied to compare the proteomes of few cultivars in common wheat, spelt and rye<sup>38,39</sup> showing that an important number of proteins were differently expressed even between common wheat and spelt, both being hexaploid species. Comparing the protein expression levels in 150 common wheat cultivars, our recent study demonstrated a large impact of the environment and of different cultivars on the expression of a range of proteins<sup>40</sup>. However, to the best of our knowledge, the flour proteomes of common wheat, spelt, durum, emmer and einkorn have yet not been compared using modern proteomics technology.

In the present work, we utilized high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS) based label-free quantitative (LFQ) proteomics to characterize the

<sup>1</sup>State Plant Breeding Institute, University of Hohenheim, Fruwirthstr. 21, 70599 Stuttgart, Germany. <sup>2</sup>Institute for Immunology and Research Center for Immune Therapy (FZI), University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany. <sup>3</sup>Institute of Translational Immunology and Research Center for Immune Therapy (FZI), University Medical Center of the Johannes Gutenberg University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany. <sup>4</sup>Division of Gastroenterology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Boston, MA 02215, USA. <sup>5</sup>These authors contributed equally: Muhammad Afzal, Malte Sielaff. <sup>6</sup>These authors jointly supervised this work: Detlef Schuppan, Stefan Tenzer, C. Friedrich H. Longin. ✉email: [friedrich.longin@uni-hohenheim.de](mailto:friedrich.longin@uni-hohenheim.de)

proteome of the whole-grain flour of ten cultivars for each of five wheat species all grown in three diverse environments. Our objectives were to (i) elaborate a high-resolution reference proteome of five wheat species, (ii) quantify the effects of the species, cultivars within species and the environment on protein abundance / expression level, and (iii) elucidate similarities and differences in the proteomes of different wheat species based on protein patterns related to allergies, immune activation and nutritional quality for improved health and wheat supply chains.

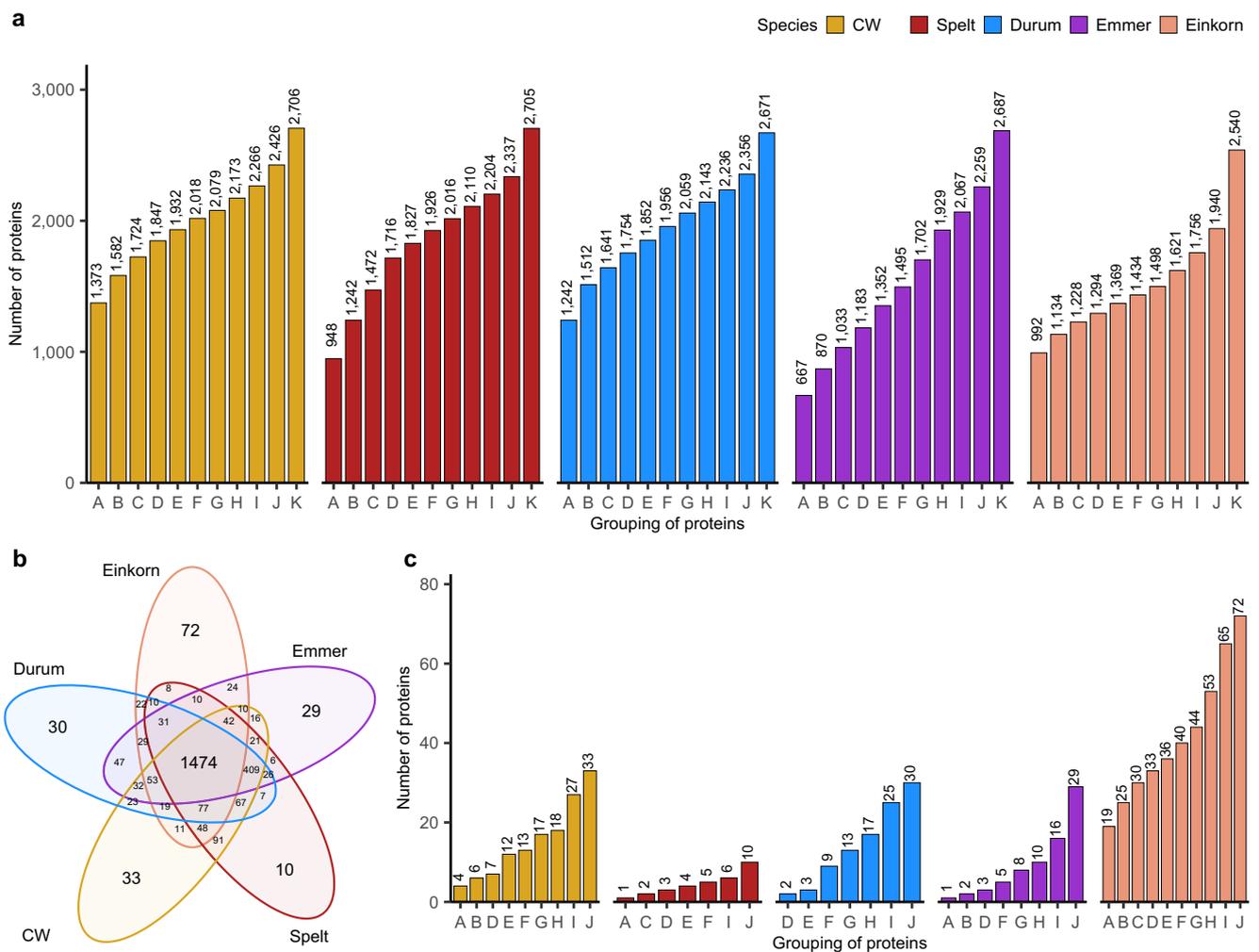
## RESULTS AND DISCUSSION

In our analysis, we identified 17,277 peptide sequences and 2,896 different proteins across 150 flour samples, representing, to our knowledge, the largest proteome study in cereals to-date. Moreover, the investigation of ten cultivars for each species grown in three diverse environments enabled the in-depth evaluation of the effects of the species, cultivars within species and the environment on the protein expression.

### Basis for future in-depth proteomic research in wheat species

Our proteomic analyses identified 2706, 2705, 2671, 2687 and 2540 proteins in common wheat, spelt, durum, emmer and einkorn, respectively (Fig. 1a). Interestingly, these numbers were quite similar between species, although the composition of the protein sequence database used for searching the MS spectra was biased towards entries from common wheat (38% of all entries) and durum wheat (51% of all entries) due to the lack of reference proteomes for some of the analyzed species. These findings were in agreement with a study comparing only the hexaploid species common wheat and spelt<sup>38</sup> indicating a high sequence homology across wheat species irrespective of different ploidy levels.

Overall, protein abundances were highly affected by choice of cultivars within species and the environment, where the cultivars were grown. For instance, from the total number of identified 2,540 proteins in einkorn (Fig. 1a letter K), only 1,940 were stably expressed across all three environments in at least one cultivar (Fig. 1a letter J): Thus, 600 proteins were present/absent only due to environmental effects. Furthermore, for the remaining proteins the mean heritability was 0.24 with only 380 proteins having a



**Fig. 1 Overview of the identified proteins in five wheat species. a** Grouping of proteins based on their presence in a certain number of cultivars across three environments. The capital letters on the x-axis denote groups of proteins. These groups correspond to the proteins that were expressed in all of the three environments in all 10 cultivars (A), in at least 9 (B), 8 (C), 7 (D), 6 (E), 5 (F), 4 (G), 3 (H), 2 (I) cultivars and in at least 1 (J) cultivar. Proteins belonging to group K were found in at least one sample. For further analysis, the proteins from group J were considered. **b** Number of proteins that are unique or common between species. **c** Grouping of the unique proteins (from Fig. 1b) based on their presence in a certain number of cultivars across three environments. The letters on x-axis have the same denotation as in Fig. 1a. CW, common wheat.

**Table 1.** Potentially interesting proteins for future research and breeding grouped into protein families.

UniProt annotation / keyword	Common wheat	Spelt	Durum	Emmer	Einkorn
Glutenin, HMW, LMW	19 (0.81)	18 (0.78)	15 (0.85)	6 (0.69)	7 (0.78)
Gliadin	23 (0.81)	16 (0.81)	12 (0.81)	2 (0.64)	5 (0.82)
Puroindoline	1 (0.82)	0	0	0	2 (0.54)
Beta-amylase	7 (0.82)	4 (0.93)	2 (0.75)	1 (0.80)	1 (0.62)
1,4-alpha-glucan branching enzyme	3 (0.66)	2 (0.62)	7 (0.74)	2 (0.57)	3 (0.64)
Sucrose synthase, Sucrose_synth	3 (0.78)	3 (0.69)	5 (0.75)	1 (0.65)	0
Alpha-amylase inhibitor, amylase inhibitor	8 (0.71)	7 (0.72)	4 (0.79)	0	0
Trypsin inhibitor	2 (0.83)	3 (0.73)	1 (0.66)	0	1 (0.59)
Serpin	4 (0.75)	4 (0.76)	9 (0.85)	5 (0.60)	1 (0.57)
Non-specific lipid-transfer protein	6 (0.80)	5 (0.89)	4 (0.74)	0	1 (0.77)
Glutathione transferase, S-transferase	2 (0.74)	1 (0.70)	4 (0.80)	2 (0.69)	0
P:defense response <sup>a</sup>	15 (0.74)	11 (0.66)	15 (0.77)	10 (0.71)	8 (0.66)
P:response to water <sup>a</sup>	8 (0.69)	4 (0.60)	4 (0.78)	4 (0.74)	0
Heat shock	5 (0.64)	2 (0.57)	4 (0.73)	5 (0.76)	3 (0.58)
P:response to heat <sup>a</sup>	10 (0.74)	5 (0.65)	17 (0.70)	7 (0.75)	3 (0.63)
Late embryogenesis abundant (LEA)	2 (0.73)	0	0	0	0

HMW high molecular weight, LMW low molecular weight.

<sup>a</sup>proteins involved in a given biological process (P:) according to the gene ontology (GO) analysis.

These proteins were filtered by applying a stringent criterion within each species, i.e., heritability >0.50, missing data ≤20%, detected in all environments in ≥50% cultivars and in at least 2 of 3 environments in ≥80% cultivars. For each species, the number of proteins within specific protein families fulfilling the filtering criteria are shown with their average heritability in brackets. For more details, see Supplementary Table1.

heritability higher than 0.5 (Supplementary Fig. 1a). The heritability quantifies the cultivar's effect on the total expression of a trait and ranges from 0 to 1. The lower the heritability, the higher the impact of the environment vs. the cultivar on a trait's expression. Consequently, the impact of the environment including soil, climatic factors and cultivation practices on protein expression is very high, which is consistent with the literature<sup>32,34,38,40</sup>. For further discussion, we disregarded proteins, which were only affected by the environment, specifically proteins not stably present across all environments in at least one cultivar of a species (Fig. 1a letter K).

Generally, the different cultivars within a species varied considerably in their protein expression, as evidenced by the presence/absence of proteins (Fig. 1a) or a large coefficient of variation across cultivars for a protein (Supplementary Fig. 1b). For instance, in einkorn 1,940 proteins were identified, which were stably present in at least one cultivar across all three environments (Fig. 1a letter J), but only 992 proteins were present in all 10 cultivars and all environments (Fig. 1a letter A). This is in line with findings of a previous study that compared common wheat and spelt<sup>38</sup>, and highlights the necessity to use representative sets of cultivars within species grown in several environments when measuring protein abundances.

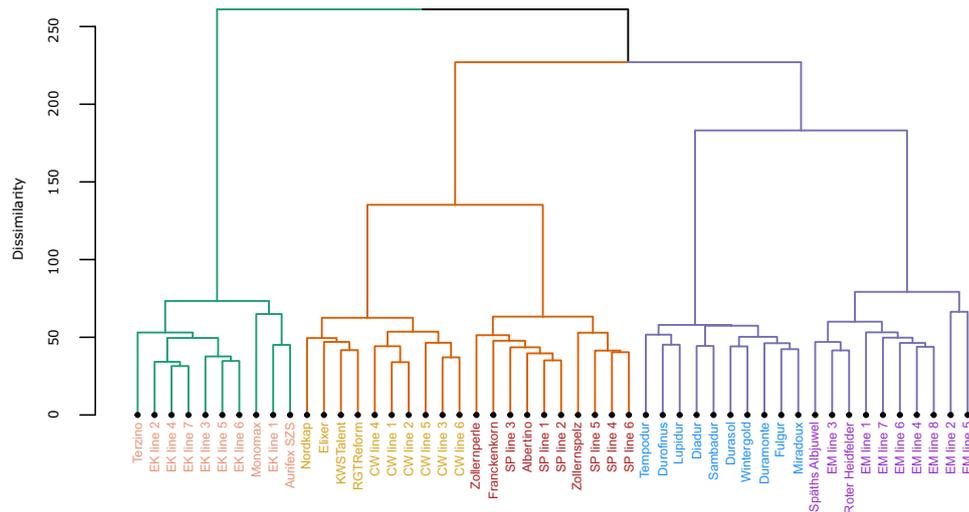
Considering the high environmental impact on protein expression, we elaborated a list of proteins mainly affected by the genetics for future research and breeding. These proteins might be successfully manipulated across future wheat supply chains by choice of cultivars. We therefore selected only proteins within each species, which (i) had a heritability >0.50, (ii) had missing data ≤20%, and (iii) were detected in all environments in ≥50% cultivars and in at least 2 of 3 environments in ≥80% cultivars. These were 845, 611, 863, 262 and 296 proteins in cultivars of common wheat, spelt, durum, emmer and einkorn, respectively (Supplementary Table 1). This list contained proteins from important families such as proteins crucial for baking quality (glutenins, gliadins), the starch pathway (beta-amylases, glucan-branching enzymes, sucrose synthases), confirmed allergens

(enzyme inhibitors, serpins, lipid transfer proteins) and the plants' response to field conditions (heat shock, heat and drought response proteins, late embryogenesis abundant proteins) and others that were partly present across different wheat species (Table 1). Moreover, many proteins on our list of "hot candidate proteins" for future wheat supply chains have still unknown or rather descriptive names warranting urgent future research. Summarizing, our present high-coverage proteomics study provides a solid basis for future in-depth research on proteome and protein functions across different wheat species.

### Five wheat species can be separated by their patterns of protein expression

The hierarchical clustering of the 50 cultivars from five wheat species using 2,774 proteins clearly separated the cultivars into five groups corresponding to the five species (Fig. 2). The clustering reflected the genetic distance between the species by depicting smaller distances between species with the same ploidy level. For instance, the two hexaploid species common wheat and spelt clustered more closely together than einkorn and common wheat, as did the two tetraploid species durum and emmer, which underlines the validity of our proteomic workflow.

This separation was further corroborated by proteins unique to an individual species or present only in few but not all species (Fig. 1b, c; Supplementary Fig. 2). For instance, the diploid einkorn had the highest number of unique proteins with ≥40 unique proteins present in at least four einkorn cultivars. By contrast, 1,474 proteins were jointly expressed across all five wheat species (Fig. 1b), with the highest (lowest) number of proteins expressed jointly in pairwise comparisons between common wheat vs. spelt (einkorn vs. emmer; Supplementary Fig. 2). However, >50% of the joint proteins between any pair of species were expressed with a statistically significantly different abundance (Fig. 3), showing a tendency that the larger the difference between the ploidy levels of the species, the higher the percentage of differentially expressed proteins. For instance, 52% of the joint proteins between spelt and common wheat (Fig. 3a) showed a



**Fig. 2 Hierarchical clustering of 50 cultivars belonging to five wheat species.** The colors of the lines (branches) highlight three clusters of cultivars corresponding to their ploidy level. The colors of the cultivar labels correspond to the color legend for species in Fig. 1a. In total, 2774 proteins were used for clustering, which included proteins unique to one of the five species and proteins common between species. All those proteins were identified across all three environments in at least one cultivar and hence belonged to the group J in Fig. 1a.

significantly different expression, which reached 78% for einkorn vs. common wheat (Fig. 3g). Thereby a higher number of proteins was downregulated than upregulated in einkorn compared to the other wheat species (Supplementary Fig. 3).

Overall, 254 proteins showed not only a statistically different expression between different pairs of species but also passed a stringent threshold of  $\pm 3 \log_2$  fold change (equivalent to an 8-fold up/downregulation) (Fig. 3 orange points, Supplementary Table 2). These proteins belonged to the important protein families mentioned above (baking quality, starch pathway, allergens, the plants' stress response) (Table 2). Except for our recent comparison between common wheat and spelt<sup>38</sup>, no such analysis has been published to-date.

Owing to the lack of reference proteomes for spelt, emmer and einkorn, our study might be biased in the way that proteins unique to these species have not been detected, which may further increase the differences among species. However, to date our study provides the highest proteome coverage regarding the identified and quantified proteins across five wheat species.

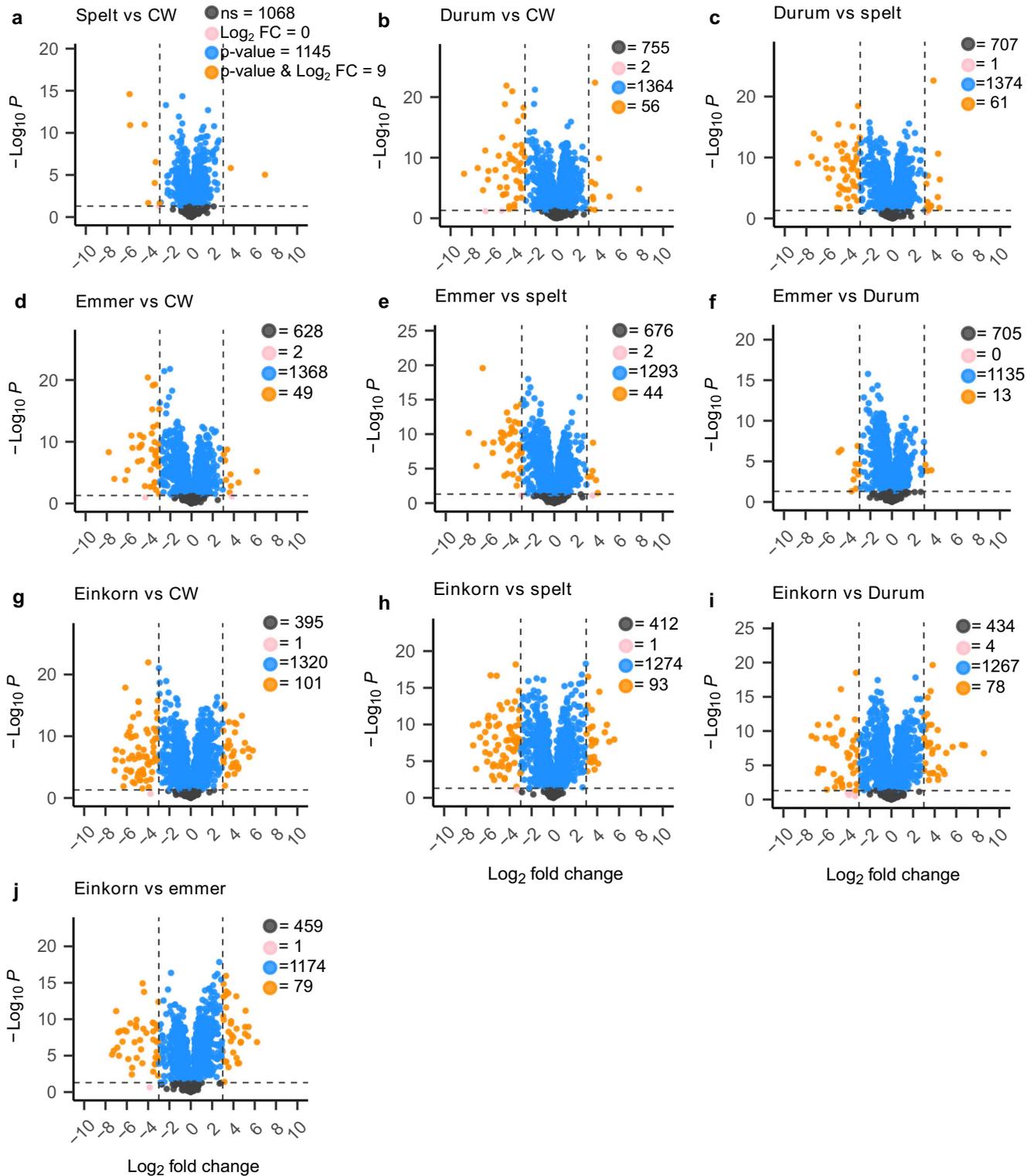
### Potential and known allergenic proteins are largely reduced in einkorn

Wheat is an important and usually healthy staple crop for human and animal nutrition, but a sizable population suffers from inflammatory wheat sensitivities. These are celiac disease, IgE-mediated and non-IgE-mediated (type 2) wheat allergy, and innate immune activation by ATI-proteins, the latter two possibly contributing to non-celiac wheat sensitivity (NCWS)<sup>14,17–23,25,30,31,41–45</sup>. As most potential allergens are proteins and large differences in the proteomes of the different wheat species were elaborated above, we investigated the distribution of potential allergenic proteins across the wheat species in more detail. We followed the approach of Zimmermann et al.<sup>39</sup> to compile a list of allergens based on the information from data about seed-borne wheat allergens<sup>27</sup> and the allergome database (<http://www.allergome.org/index.php>)<sup>46</sup>, and additionally included ATIs.

The sum of all these potential allergenic proteins was clearly different among the species and corresponded almost perfectly with the ploidy levels (Fig. 4a). While their total abundance was similar in hexaploid common wheat and spelt, they were roughly reduced two-fold in tetraploid durum and emmer, and 5.4-fold in diploid einkorn. These differences were due to both the different

numbers of potential allergens and different protein abundances. By contrast, total grain protein content (GPC) was slightly higher in einkorn, spelt and emmer compared to common wheat (Supplementary Fig. 4), which is in line with prior findings<sup>32,47</sup>, with differences in GPC of other species compared to common wheat ranging between 1% and 20%. In common wheat, spelt and durum, almost half of the potential allergenic proteins had a heritability  $>0.5$ , and across all species their coefficient of variation with heritability  $>0.5$  ranged between 7% and 261% (Fig. 4a). Therefore, the abundance of these potential allergens can be reduced in a targeted way using proteomics to monitor breeding and cultivar choice, which confirms recent findings on gluten and ATI composition across different wheat cultivars<sup>33,34,48</sup>. This would, however, require the development of rapid test methods, which can be used in daily business across wheat supply chains. Our reference proteome can be used as starting point, e.g., by concentrating on allergens with high heritability and coefficient of variation across cultivars within a species.

Besides the quantity of allergens, their distribution within species also varied considerably (Fig. 4b). While the allergens of common wheat were largely represented by ATIs, gliadins, HMW and LMW glutenins, more than 50% of the einkorn allergens (substantially lower abundances than in other wheat species, Fig. 4a) were comprised of gliadins and LMW glutenins. Consistent with the published literature<sup>33,48</sup>, einkorn had significantly lower quantity of ATIs compared to the other wheat species i.e., 7.2, 7.3, 5.2 and 4.7-fold lower than in common wheat, spelt, durum and emmer, respectively (Fig. 4c). The ATIs are implicated as major allergens and also as activators of TLR4 in animal models of diseases<sup>17–19,21–23,33,48</sup>. Einkorn ATIs were mainly attributable to CMX1/CMX3 (UniProt accession M8A1S2) (Fig. 4d). Interestingly, in einkorn the predominant ATIs of type CMX1/CMX2/CMX3 have not been described to inhibit amylase activity<sup>49</sup> and CMX1/CMX3 are neither listed as seed-borne wheat allergens<sup>27</sup> nor in the allergome database (<http://www.allergome.org/index.php>)<sup>46</sup>. Furthermore, another ATI (UniProt accession C5J3R4; Description, Trypsin inhibitor OS = Triticum monococcum subsp. monococcum OX = 408188 GN = Eti-Am1) was found only in emmer and einkorn, confirming our recent analyses focused on ATIs<sup>35</sup>. By contrast, ATIs of hexaploid common wheat and spelt, were mainly represented by 0.19, CM1, CM2, CM3 and CM16, whereas in tetraploid durum and emmer the ATIs were mainly 0.53, CM3 and CM16 (Fig. 4d). Iacomino et al.<sup>36</sup> showed that the einkorn ATIs



**Fig. 3** Results of the comparisons for individual proteins between different pairs of wheat species. **a–j** Volcano plots of proteins that were common between different pairs of wheat species. The x-axis shows the log<sub>2</sub>-scale fold change (Log<sub>2</sub>FC) in the LFQ abundance of proteins between two species. The y-axis shows the negative log<sub>10</sub>-scale *p* value from the t-test. The vertical dashed lines denote  $\pm 3$  Log<sub>2</sub>FC threshold, while the horizontal dashed line denotes *p* value threshold of 0.05, e.g., the orange points in **(a)** show proteins, which are downregulated (Log<sub>2</sub>FC < -3) or upregulated (Log<sub>2</sub>FC > 3) in spelt compared to common wheat and passed both thresholds i.e., *p* value and Log<sub>2</sub>FC. Color of the points represents significance status of proteins for *p* value (<0.05) and Log<sub>2</sub>FC ( $\pm 3$ ): gray, non-significant for both thresholds; pink, significant Log<sub>2</sub>FC only; blue, significant *p* value only; orange, significant for both thresholds. CW, common wheat.

**Table 2.** Proteins with potentially important functions, which were significantly differently expressed (t-test, *p* value <0.05) between species.

UniProt annotation / keyword	NPS	GO annotations <sup>a</sup>	NPU				
			Cw	Sp	Du	Em	Ei
Glutenin, HMW, LMW	50	F:nutrient reservoir activity	2	–	1	1	3
Gliadin	41	F:nutrient reservoir activity	6	2	2	–	4
Puroindoline	7	F:nutrient reservoir activity	–	–	–	–	–
Beta-amylase	7	F:beta-amylase activity	1	–	–	1	–
1,4-alpha-glucan branching enzyme	12	P:glycogen biosynthetic process	1	–	1	–	–
Starch synthase	12	11 F:glycogen (starch) synthase activity	–	–	1	–	1
Sucrose synthase, Sucrose_synth	8	F:sucrose synthase activity	–	–	–	–	–
Alpha-amylase inhibitor, amylase inhibitor, Trypsin inhibitor, BBTI, CTI	30	F:serine-type endopeptidase inhibitor activity	1	–	–	–	1
Serpin	16	F:serine-type endopeptidase inhibitor activity	1	–	–	–	–
Nonspecific lipid-transfer protein	10	P:lipid transport	–	–	–	–	–
ATP synthase	14	P:ATP synthesis coupled proton transport	–	–	–	–	–
Glutathione transferase, S-transferase	8	7 P:glutathione metabolic process, 1 F:glutathione dehydrogenase (ascorbate) activity	–	–	–	–	–
Defensin, Barwin, Thaumatin-like, Thionin, Knot1	13	P:defense response	–	–	1	–	2
Dehydrin	4	P:response to water	–	1	–	–	–
Heat shock, SHSP	42	F:ATP hydrolysis activity, P:response to heat	4	–	7	3	6
Late embryogenesis abundant, LEA	4	–	–	–	–	–	–

*NPS* number of proteins, which were significantly differently expressed (t-test, *p* value <0.05) between species, *GO* gene ontology, *NPU* number of proteins, which were unique to only one of the five species, *Cw* common wheat, *Sp* spelt, *Du* durum, *Em* emmer, *Ei* einkorn, *HMW* high molecular weight, *LMW* low molecular weight, *BBTI* Bowman-Birk type trypsin inhibitor, *CTI* chymotrypsin inhibitor.  
For more details, see Supplementary Table 2.  
<sup>a</sup>GO annotations provided as either molecular function (F:) or biological process (P:) obtained from the Gene Ontology (GO) analysis.

were more susceptible to in vitro enzymatic hydrolysis than the fairly pepsin-trypsin resistant ATIs of tetra- and hexaploid wheats<sup>19,21–23,50</sup>, and therefore largely proteolytically degraded during food processing and especially upper gastrointestinal passage, resulting in absent or reduced ability to trigger innate immunity<sup>36</sup>. Similarly, Sievers et al.<sup>51</sup> showed that einkorn could be beneficial for people who are sensitive only to ATIs but might not be safe for individuals suffering from wheat allergy.

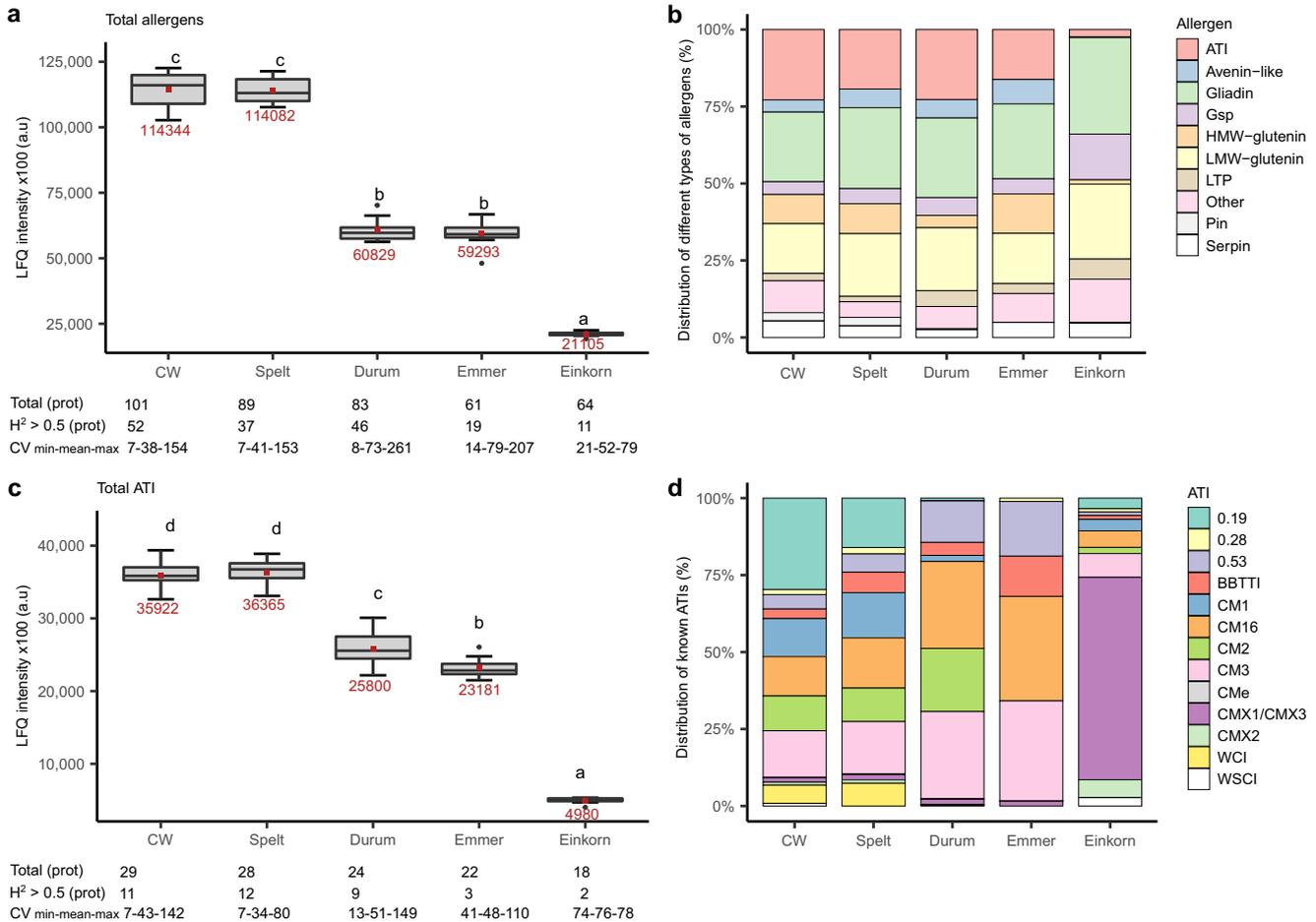
Summarizing, our study demonstrates a much lower abundance of potential allergens and ATIs in einkorn, and also in durum and emmer in comparison with hexaploid common wheat and spelt. However, allergen databases appear to be biased towards having more allergens from common wheat than from other wheat species probably due to the limited use and cultivation of alternative wheat species such as spelt, emmer and einkorn. Furthermore, the utilized proteome reference included only the reference proteomes of *T. turgidum* ssp. *durum* (51% of all database entries), *T. aestivum* ssp. *aestivum* (38% of all database entries) and *T. urartu* (9% of all database entries). Consequently, our approach might have fallen short to identify all potential allergens from einkorn, emmer, durum and spelt. Nevertheless, we speculate that emmer, durum and einkorn still have a considerably lower number and lower abundance of allergenic proteins, because the identified differences are very large and sequence homology was high so that high number of proteins could be identified in all wheat species regardless of the limitations mentioned above.

The need to find alternative cereal crops with reduced allergenicity or ATIs is highlighted by the increasing number of preclinical and especially clinical studies comparing different wheats. First explorative preclinical studies suggest a potential better tolerability of ancient vs. modern wheat cultivars for patients with wheat allergy and NCWS. In these studies, ancient diploid and also tetraploid wheats like einkorn and emmer were

indeed better tolerated by NCWS patients, many of whom had either an IgE positive or an IgE negative (type 2) wheat allergy, when compared to modern hexaploid wheats<sup>18,52–58</sup>. Similarly, Picascia et al.<sup>59</sup> investigated the effect of the diet prepared from einkorn and common wheat flours on the immune response in celiac disease patients and concluded that einkorn caused a lower in vivo T-cell response in comparison with common wheat. However, these studies were not based on representative samples including different cultivars from different wheat species grown at comparable environmental conditions, all of them largely influencing the proteomic profiles of flour samples as indicated in this study. Consequently, better targeted controlled clinical trials using wheats with a defined low content of potential allergens are urgently needed and can be started based on the reference proteomes delivered by current study. Besides einkorn and other wheat cultivars with low abundance of potential allergens, the effect of different flour processing and bread making procedures, such as long sourdough fermentation, on the abundance and activity of allergens should also be investigated in pursuit of identifying healthier wheat products especially for people with wheat related disorders.

#### **Outlook: Einkorn as sustainable crop for marginal environments, but potential health benefits have to be urgently validated**

In addition to the lower amount of potential allergens with yet lacking clinical proof described above, einkorn contains more protein and considerably higher amounts of valuable trace compounds compared with common wheat, such as vitamin E, luteins, steryl ferulates<sup>60,61</sup>, minerals like Fe and Zn<sup>62</sup> in addition to several other minerals such as Ca, Cu, K, Mg, Mn, P, and S<sup>63</sup> – all these compounds being important for a healthy diet<sup>61,62</sup>. However, the bioavailability of these compounds in breads and



**Fig. 4 Allergenic proteins in wheat species.** LFO intensity/abundance of different families of proteins per cultivar in different wheat species for (a) potential allergens, and (c) proteins identified as ATIs. Boxes with different letters indicate significant differences between the wheat species (ANOVA with Tukey's test,  $p < 0.05$ ). Each box summarizes the content of 30 samples (ten cultivars grown at three locations). Mean values for each species are highlighted in red color. Furthermore, within each species the distribution of (b) allergens, and (d) ATIs with a known type is shown. CW, common wheat; Total (prot), number of allergenic proteins or ATIs used to calculate total LFO intensity (%); H<sup>2</sup> > 0.5 (prot), number of allergenic proteins or ATIs with heritability >0.5; H<sup>2</sup>, heritability; CV<sub>min-mean-max</sub>, minimum-mean-maximum values of the coefficient of variation (%) for allergenic proteins or ATIs with heritability >0.5; LFO, label-free quantification; ATI, amylase/trypsin inhibitor; Gsp, grain softness protein; HMW, high molecular weight; LMW, low molecular weight; LTP, lipid transfer protein; Pin, puroindoline; BBTTI, Bowman-Birk type trypsin inhibitor; CM, chloroform/methanol soluble protein; WCI, wheat chymotrypsin inhibitor; WSCI, wheat subtilisin-chymotrypsin inhibitor.

other cereal products and their impact on human health have yet not been investigated warranting urgent further research.

Agronomically, einkorn shows a considerably higher protein yield efficiency than common wheat<sup>32</sup>, almost complete resistance against fungi<sup>8</sup> and flexibility to sow the same cultivar before or after winter, which does not exist in other cereals. However, compared to common wheat, einkorn plants are taller and, thus, more prone to lodging than common wheat. Furthermore, einkorn has almost 70% less grain yield under good soil conditions than common wheat<sup>47</sup>. Owing to the necessity of increasing agricultural productivity per available cultivated land to feed an increasing world population, einkorn cannot replace widely cultivated common wheat. Nevertheless, in marginal environments the productivity of common wheat is markedly reduced<sup>64</sup>, whereas einkorn performs well<sup>65</sup>. The marginal environments include sandy soils and higher altitudes in mountainous regions and/or lack of nitrogen fertilizer due to high costs, environmental restrictions or organic farming. Consequently, considering the lowest allergen and ATI contents and the high amounts of nutritious ingredients of einkorn, clinical trials are urgently needed to validate these potential health benefits, as einkorn could be a

promising sustainable alternative crop for marginal regions enhancing agro-biodiversity.

## METHODS

### Plant material and field trials

We investigated 10 cultivars of each of the five species of wheat namely common wheat (*T. aestivum* ssp. *aestivum*,  $2n = 6 \times = 42$ , A<sup>4</sup>A<sup>4</sup>BBDD), spelt (*T. aestivum* ssp. *spelta*,  $2n = 6 \times = 42$ , A<sup>4</sup>A<sup>4</sup>BBDD), durum (*T. turgidum* ssp. *durum*,  $2n = 4 \times = 28$ , A<sup>4</sup>A<sup>4</sup>BB), emmer (*T. turgidum* ssp. *dicoccum*,  $2n = 4 \times = 28$ , A<sup>4</sup>A<sup>4</sup>BB), and einkorn (*T. monococcum* ssp. *monococcum*,  $2n = 2 \times = 14$ , A<sup>m</sup>A<sup>m</sup>). The selected cultivars of each wheat species include very important cultivars representing the recent market in Germany. In addition, the best and latest breeding lines available from multiple environment field trials were added to the selected cultivars (Supplementary Table 3).

The field trials were conducted as winter cropping, i.e., sowing in October 2018 and harvest in July 2019, at three diverse locations for each species in Germany/Austria. The trial locations with their GPS coordinates for each species are provided within

brackets followed by name of the species, such as, common wheat (DSV-Leutewitz - 51°8'58"N, 13°21'46.908"E; Eckartsweier - 48°31'45"N, 7°51'18"E; Stuttgart-Hohenheim - 48°42'50"N, 9°12'58"E), spelt (Eckartsweier - 48°31'45"N, 7°51'18"E; Stuttgart-Hohenheim - 48°42'50"N, 9°12'58"E; Oberer Lindenhof - 48°28'26"N, 9°18'12"E), durum (Eckartsweier - 48°31'45"N, 7°51'18"E; Stuttgart-Hohenheim - 48°42'50"N, 9°12'58"E; Probstdorf - 48°10'19"N, 16°37'13"E), emmer (Stuttgart-Hohenheim - 48°42'50"N, 9°12'58"E; Ihingerhof - 48°44'44"N, 8°55'23"E; Oberer Lindenhof - 48°28'26"N, 9°18'12"E), einkorn (Eckartsweier - 48°31'45"N, 7°51'18"E; Ihingerhof - 48°44'44"N, 8°55'23"E; Oberer Lindenhof - 48°28'26"N, 9°18'12"E). The wheat species were investigated in separate but adjacent trials at individual locations using an un-replicated field design randomized separately for each species across test locations. All trials received the same field treatments of intensive conventional farming practices except for nitrogen (N) fertilization, where common wheat, durum, spelt, emmer and einkorn were fertilized to reach the following level of N including N measured in the soil ("Nmin"): 180, 180, 160, 60 and 60 kg N/ha, respectively. This individual adjustment was done to reflect recent agricultural practice in conventional production. Furthermore, owing to its field resistance to fungal diseases einkorn did not receive any fungicide treatment in contrast to all other species. Field net plot size was 5 m<sup>2</sup> in all locations. All plots were machine-sown and combine-harvested. All samples of spelt, emmer and einkorn were dehulled and cleaned using a Mini-Petkus seed cleaner (Röber, Bad Oeynhausen, Germany) to separate hulls, straw and damaged kernels. Dehulling was performed using a classical stone mill, in which the stone was replaced by hard rubber. For common wheat, seed cleaning was also performed using the Mini-Petkus seed cleaner in order to remove chaff and straw particles, which were still present after combine harvesting.

### Laboratory analyses

Three observations (samples) from three diverse locations were used to calculate the mean abundance of a protein per cultivar of each species. However, for lab analysis we used one technical replicate for each sample. While the use of more technical replicates is preferable, we had to compromise due to the large number of samples to analyze under a given budget and time-limit. This approach is justified, since from numerous studies on different traits in field trials, it is well known to and accepted by the scientific community that a large variance in data arises due to differences in the conditions between different growing locations. Given this location-dependent variability, the data quality improves and becomes more representative for general statements about the expression of a trait, such as proteins in different species, if the number of locations is increased at the expense of the number of technical replicates and not vice versa. Furthermore, the low number of technical replicates was accounted for during the statistical analysis to estimate the mean values across three locations by adjusting for field trial and effects of lab analyses.

**Protein extraction.** From each cultivar of common wheat, spelt, durum, emmer and einkorn 20 mg of whole-grain flour were weighed into 1.5 mL plastic tubes (Protein LoBind tubes, Eppendorf, Hamburg, Germany). Next, 50 µL of LC-MS grade water were added and the tubes were vortexed until the flour was completely resuspended. Immediately afterwards, 950 µL of extraction buffer composed of 7 M urea, 2 M thiourea, 2% (w/v) CHAPS, 5 mM dithiothreitol (DTT) and LC-MS grade water were added and the tubes were vortexed again. After incubation at 22 °C for 10 min, the samples were centrifuged at 16,000 x g and 22 °C for 10 min. The clear upper layer of the supernatants was used for all subsequent steps. Average protein concentrations

were determined from extracts of species-specific flour mixtures using the Pierce 660 nm protein assay (Thermo Scientific, Rockford, IL, USA; Supplementary Fig. 4).

**Sample preparation.** Proteins were purified and digested into LC-MS-compatible tryptic peptides using a filter-assisted sample preparation protocol (FASP)<sup>66,67</sup> as described before<sup>68</sup> with minor modifications. Briefly, 30 µg of protein extracts were loaded onto centrifugal ultrafiltration devices (Nanosep with 30 K MWCO Omega membrane 30 K MWCO, Pall, Port Washington, NY, USA) and centrifuged at 16,000 x g and 22 °C for 15–30 min until the liquid completely passed through the membrane. Disulfide bonds were reduced using DTT followed by alkylation of free cysteines using iodoacetamide (IAA). The alkylation reaction was quenched by the addition of DTT. After each step, the membrane was washed once using a buffer containing 8 M urea and 100 mM Tris-HCl (pH 8.5). Finally, buffer exchange was performed washing the membrane three times with a buffer containing 50 mM ammonium bicarbonate and LC-MS grade water. On-filter tryptic digestion was performed by the addition of trypsin (Trypsin Gold, Promega, Madison, WI, USA) at a protease-to-protein ratio of 1:50 (w/w) and overnight incubation at 37 °C. Tryptic peptides were collected into fresh tubes by centrifugation and an additional wash of the membrane using 50 mM ammonium bicarbonate. The flow through was acidified adding trifluoroacetic acid (TFA) to achieve final concentration of 0.5% (v/v). Peptides were loaded onto Sep-Pak tC18 96-well cartridges (Waters Corporation, Milford, MA, USA) and desalted using 0.1% (v/v) TFA in LC-MS grade water as wash solvent and 0.1% (v/v) TFA in 50% (v/v) acetonitrile/water as elution solvent. Purified peptides were lyophilized and reconstituted in 20 µL of 0.1% (v/v) formic acid (FA) in LC-MS grade water prior to LC-MS analysis.

**Liquid chromatography-mass spectrometry.** Tryptic peptides of each sample were sequentially analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a nanoACQUITY UPLC system (Waters Corporation) coupled to a SYNAPT G2-S mass spectrometer (Waters Corporation) via a NanoLockSpray dual electrospray ionization source (Waters Corporation). Microflow LC and source interface were set up as described before<sup>69</sup>. Peptides in amounts of 1.5 µL were loaded onto an HSS-T3 C18 reversed phase column (Waters Corporation) with a length of 250 mm and an inner diameter of 300 µm and separated by gradient elution using LC flow rates of 5 µL/min and 60 min LC methods. Precursor and fragment ion mass spectra were recorded using an ion mobility-enhanced data-independent acquisition strategy as described before (UDMS<sup>E</sup>)<sup>68</sup>.

**Data processing and label-free quantification.** Raw UDMS<sup>E</sup> data were processed using ProteinLynx Global Server v3.0.2 (PLGS, Waters Corporation) and searched against a database compiled of proteins of the genus *Triticum* (UniProtKB release 2020\_05, taxon ID: 4564, 367,831 entries) including the reference proteomes of *T. turgidum* ssp. *durum* (51% of all database entries), *T. aestivum* ssp. *aestivum* (38% of all database entries) and *T. urartu* (9% of all database entries) plus 171 common MS contaminants using following parameters: Trypsin was specified as digestion enzyme, two missed cleavages per peptide were allowed for initial database search, carbamidomethylation of cysteines was set as fixed, and methionine oxidation as variable modification. The false discovery rate (FDR) was calculated in PLGS by searching a database of reversed protein sequences and a cutoff of 0.01 was applied.

Label-free quantification (LFQ) including retention time alignment, feature clustering, cross-run normalization and protein inference was performed using ISOQuant v1.8<sup>68</sup>. Only peptides without missed cleavages, a minimum sequence length of seven amino acids, a minimum PLGS score of 6.0 and no variable

modification were considered for quantification. An FDR cutoff of 0.01 was applied at the peptide and protein level in ISOQuant, ensuring a 1% FDR on dataset level. Proteins identified by at least two different peptides were quantified by averaging the intensities of the three peptides with the highest intensities belonging to the respective protein (Top3 method)<sup>70</sup>. Abundances of shared peptides were redistributed between proteins based on relative abundances of uniquely assigned peptides (see ISOQuant manual for details; <http://www.immunologie.uni-mainz.de/isoquant/index.php?slab=user-manual#x1-40006.7>). TOP3-based quantification provides an estimate of the total amount of each protein in a sample. By summing up over all detected and quantified proteins, the relative amount of each protein in the respective proteome (i.e. parts per million of total protein) can be determined. This value is independent of the total amount of protein in the sample or on column.

**Database search and homology filtering - detailed procedure.** During PLGS database search, each detected peptide is mapped to all proteins in the database containing the respective peptide. This is initially performed on a run-by-run basis in PLGS. Subsequently, during the data processing in ISOQuant, protein groups are filtered on entire dataset basis (i.e. taking peptide and protein information from all runs, filtering at 1% FDR on peptide level initially and taking only peptides without missed cleavages, a minimum sequence length of seven amino acids, a minimum PLGS score of 6.0 and no variable modification), then based on the Occam's Razor principle, resulting in a reduced protein list (also filtered at 1% FDR) that can explain all peptides passing the above criteria in the dataset.

**Blast2GO analysis.** Automatic functional annotation of proteins identified by LC-MS/MS was performed using the software tool Blast2GO v5.2.5<sup>71</sup>. Blast2GO uses the BLAST algorithm to identify similar proteins and transfers already existing Gene Ontology (GO) annotations to the queried protein sequences. In addition, InterProScan is used to obtain protein family and domain information which are converted and merged to GOs. Blast searches were performed against the NCBI database of non-redundant protein sequences (nr) using the blastp algorithm. Otherwise, the Blast2GO workflow (Blast, InterProScan, mapping and annotation) was carried out with standard parameters.

### Phenotypic data analysis

Phenotypic data analysis was performed separately for each species according to the linear mixed model, given in Eq. (1):

$$y_{ik} = \mu + v_i + env_k + e_{ik}, \quad (1)$$

where  $y_{ik}$  is the phenotypic (= measured) observation for the  $i$ th cultivar tested in the  $k$ th environment,  $\mu$  is the general mean,  $v_i$  the effect of the  $i$ th cultivar,  $env_k$  the effect of the  $k$ th environment, and  $e_{ik}$  is the residual error. Variance components, which are variances due to cultivars, environments and residual error, were estimated using the restricted maximum likelihood (REML) method assuming a random model in a classical one-stage analysis<sup>72</sup>. A likelihood ratio test with model comparisons was performed<sup>73</sup> to check for significance of the variance components. Average values of the proteins across the different environments were determined as best linear unbiased estimates (BLUEs) assuming fixed genetic (cultivar) effects. Heritability estimates ( $h^2$ ) were computed following Piepho and Möhring<sup>74</sup> as given in Eq. (2):

$$h^2 = 1 - \frac{\vartheta}{2\sigma_G^2}, \quad (2)$$

where  $\vartheta$  is the mean variance of a difference of two best linear unbiased predictors and  $\sigma_G^2$  the genotypic variance (cultivar

variance). All analyses were performed utilizing the statistical software R<sup>75</sup> and ASReml 3.0<sup>76</sup>.

### Comparison of the proteomes of different wheat species

For the t-test, BLUEs of proteins per cultivar in each species were used. We implemented Student's t-test ( $\alpha = 0.05$ )<sup>77</sup> to compare the abundance of proteins between each pair of wheat species. For the t-test, assumption of the equality of variances between groups was examined by applying Levene's test<sup>78</sup>. If Levene's test was significant ( $p < 0.05$ , meaning that the variances between groups were not equal), then the more robust Welch's t-test was conducted instead of the regular t-test, with corrected degrees of freedom. Student's t-test and Levene's test were implemented using the statistical software R<sup>75</sup>.

Volcano plots were generated for each pair of species to identify proteins, which abundance was statistically significantly different (t-test,  $p < 0.05$ ) and was above/below an arbitrary threshold of  $\pm 3 \log_2$  fold change ( $\log_2FC$ ). R-package *Enhanced-Volcano* was used to produce volcano plots.  $\log_2FC$  was calculated using the formula given in Eq. (3):

$$\log_2FC_p = u_i - u_j, \quad (3)$$

where  $\log_2FC_p$  is the abundance of protein  $p$  in species  $i$  relative to species  $j$ ,  $u_i$  and  $u_j$  are the  $\log_2$  mean abundance of protein  $p$  in species  $i$  and species  $j$ , respectively.

For hierarchical clustering the data were scaled and Euclidian distance was calculated. Hierarchical clustering was performed using "hclust" function in statistical software R<sup>75</sup> by implementing Ward's method<sup>79</sup>.

### Identification of allergenic proteins

We used the list of seed-borne wheat allergens<sup>27</sup> curated based on the databases of allergenic protein families, AllFam ([www.meduniwien.ac.at/allfam](http://www.meduniwien.ac.at/allfam)) and AllergenOnline ([www.allergenonline.org](http://www.allergenonline.org)) to identify potential allergenic proteins in flour samples from wheat species in the current study. We further extended the identification of allergenic proteins using the comprehensive allergome database (<http://www.allergome.org/index.php>)<sup>46</sup>, which contains identified, characterized and peer-reviewed allergenic proteins including proteins from the official Allergen Nomenclature of World Health Organization (WHO) and International Union of Immunological Societies (IUIS) (<http://www.allergen.org/index.php>). The UniProt accessions were used to map proteins to the allergome database through the frontend browser of the UniProt database (<https://www.uniprot.org/>) and the corresponding Allergome IDs were retrieved (Supplementary Table 4). In addition to the identification of allergenic proteins using the aforementioned databases, we used the protein annotations from the UniProt database to search for amylase/trypsin inhibitors (ATIs) among proteins quantified in this study. The list of ATIs is provided in Supplementary Table 4.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

The mass spectrometry proteomics data, including raw files, peptide and protein quantification reports, have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository<sup>80</sup> with the dataset identifier PXD028676.

Received: 29 November 2022; Accepted: 6 March 2023;  
Published online: 25 March 2023

## REFERENCES

- Weegels, P. L. The Future of Bread in View of its Contribution to Nutrient Intake as a Starchy Staple Food. *Plant Foods Hum. Nutr.* **74**, 1–9 (2019).
- Beres, B. L. et al. A Systematic Review of Durum Wheat: Enhancing Production Systems by Exploring Genotype, Environment, and Management (G × E × M) Synergies. *Front. Plant Sci.* **11**; <https://doi.org/10.3389/fpls.2020.568657> (2020).
- Lev-Yadun, S., Gopher, A. & Abbo, S. The cradle of agriculture. *Science* **288**, 1602–1603 (2000).
- Zaharieva, M., Ayana, N. G., Hakimi, A. A., Misra, S. C. & Monneveux, P. Cultivated emmer wheat (*Triticum dicoccon* Schrank), an old crop with promising future: a review. *Genet. Resour. Crop Evol.* **57**, 937–962 (2010).
- Zaharieva, M. & Monneveux, P. Cultivated einkorn wheat (*Triticum monococcum* L. subsp. *monococcum*): the long life of a founder crop of agriculture. *Genet. Resour. Crop Evol.* **61**, 677–706 (2014).
- Hidalgo, A. & Brandolini, A. Nutritional properties of einkorn wheat (*Triticum monococcum* L.). *J. Sci. Food Agric.* **94**, 601–612 (2014).
- Longin, C. F. H. & Würschum, T. Back to the Future – Tapping into Ancient Grains for Food Diversity. *Trends Plant Sci.* **21**, 731–737 (2016).
- Miedaner, T. & Longin, C. F. H. *Neglected cereals. From ancient grains to superfood* (Erling, Glentze, 2017).
- Shewry, P. R. Wheat. *J. Exp. Bot.* **60**, 1537–1553 (2009).
- Siddiqi, R. A., Singh, T. P., Rani, M., Sogi, D. S. & Bhat, M. A. Diversity in Grain, Flour, Amino Acid Composition, Protein Profiling, and Proportion of Total Flour Proteins of Different Wheat Cultivars of North India. *Front. Nutr.* **7**, 141 (2020).
- Veraverbeke, W. S. & Delcour, J. A. Wheat Protein Composition and Properties of Wheat Glutenin in Relation to Breadmaking Functionality. *Crit. Rev. Food Sci. Nutr.* **42**, 179–208 (2002).
- Shewry, P. R. & Halford, N. G. Cereal seed storage proteins: structures, properties and role in grain utilization. *J. Exp. Bot.* **53**, 947–958 (2002).
- Shewry, P. R. What Is Gluten-Why Is It Special? *Front. Nutr.* **6**, 101 (2019).
- Catassi, C. et al. The Overlapping Area of Non-Celiac Gluten Sensitivity (NCGS) and Wheat-Sensitive Irritable Bowel Syndrome (IBS): An Update. *Nutrients* **9**, 1268 (2017).
- Schuppan, D., Junker, Y. & Barisani, D. Celiac Disease: From Pathogenesis to Novel Therapies. *Gastroenterology* **137**, 1912–1933 (2009).
- Sollid, L. M. et al. Update 2020: nomenclature and listing of celiac disease–relevant gluten epitopes recognized by CD4 + T cells. *Immunogenetics* **72**, 85–88 (2020).
- Ashfaq-Khan, M. et al. Dietary wheat amylase trypsin inhibitors promote features of murine non-alcoholic fatty liver disease. *Sci. Rep.* **9**, 1–14 (2019).
- Bellinghausen, I. et al. Wheat amylase-trypsin inhibitors exacerbate intestinal and airway allergic immune responses in humanized mice. *J. Allergy Clin. Immunol.* **143**, 201–212.e4 (2019).
- Junker, Y. et al. Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. *J. Exp. Med.* **209**, 2395–2408 (2012).
- Schuppan, D. & Zevallos, V. Wheat amylase trypsin inhibitors as nutritional activators of innate immunity. *Dig. Dis.* **33**, 260–263 (2015).
- Zevallos, V. F. et al. Nutritional Wheat Amylase-Trypsin Inhibitors Promote Intestinal Inflammation via Activation of Myeloid Cells. *Gastroenterology* **152**, 1100–1113.e12 (2017).
- Zevallos, V. F. et al. Dietary wheat amylase trypsin inhibitors exacerbate murine allergic airway inflammation. *Eur. J. Nutr.* **58**, 1507–1514 (2019).
- Caminero, A. et al. Lactobacilli Degrade Wheat Amylase Trypsin Inhibitors to Reduce Intestinal Dysfunction Induced by Immunogenic Wheat Proteins. *Gastroenterology* **156**, 2266–2280 (2019).
- Dahl, S. W., Rasmussen, S. K. & Hejgaard, J. Heterologous Expression of Three Plant Serpins with Distinct Inhibitory Specificities\*. *J. Biol. Chem.* **271**, 25083–25088 (1996).
- Fasano, A., Sapone, A., Zevallos, V. & Schuppan, D. Nonceliac gluten sensitivity. *Gastroenterology* **148**, 1195–1204 (2015).
- Salcedo, G., Quirce, S. & Diaz-Perales, A. Wheat allergens associated with Baker's asthma. *J. Investig. Allergol. Clin. Immunol.* **21**, 81–92 (2011).
- Juhász, A. et al. Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Sci. Adv.* **4**, eaar8602 (2018).
- Mameri, H. et al. Molecular and immunological characterization of wheat serpin (Tri a 33). *Mol. Nutr. Food Res.* **56**, 1874–1883 (2012).
- van Winkle, R. C. & Chang, C. The Biochemical Basis and Clinical Evidence of Food Allergy Due to Lipid Transfer Proteins: A Comprehensive Review. *Clin. Rev. Allerg. Immunol.* **46**, 211–224 (2014).
- Fritscher-Ravens, A. et al. Many Patients With Irritable Bowel Syndrome Have Atypical Food Allergies Not Associated With Immunoglobulin E. *Gastroenterology* **157**, 109–118 (2019).
- Fritscher-Ravens, A. et al. Confocal Endomicroscopy Shows Food-Associated Changes in the Intestinal Mucosa of Patients With Irritable Bowel Syndrome. *Gastroenterology* **147**, 1012–1020 (2014).
- Geisslitz, S., Longin, C. F. H., Scherf, K. A. & Koehler, P. Comparative study on gluten protein composition of ancient (einkorn, emmer and spelt) and modern wheat species (durum and common wheat). *Foods* **8**, 409 (2019).
- Geisslitz, S., Longin, C. F. H., Koehler, P. & Scherf, K. A. Comparative quantitative LC-MS/MS analysis of 13 amylase/trypsin inhibitors in ancient and modern Triticum species. *Sci. Rep.* **10**, 14570 (2020).
- El Hassouni, K. et al. Genetic architecture underlying the expression of eight  $\alpha$ -amylase trypsin inhibitors. *Theor. Appl. Genet.* **134**, 3427–3441 (2021).
- Sielaff, M. et al. Hybrid QconCAT-Based Targeted Absolute and Data-Independent Acquisition-Based Label-Free Quantification Enables In-Depth Proteomic Characterization of Wheat Amylase/Trypsin Inhibitor Extracts. *J. Proteome Res.* **20**, 1544–1557 (2021).
- Iacomino, G. et al. Triticum monococcum amylase trypsin inhibitors possess a reduced potential to elicit innate immune response in celiac patients compared to Triticum aestivum. *Food Res. Int.* **145**, 110386 (2021).
- Khodabocus, I., Li, Q., Mehta, D. & Uhrig, R. G. A Road Map for Undertaking Quantitative Proteomics in Plants: New Opportunities for Cereal Crops. In *Accelerated Breeding of Cereal Crops*, edited by A. Bilichak & J. D. Laurie (Springer US, New York, NY, 2022), pp. 269–292.
- Afzal, M. et al. High-resolution proteomics reveals differences in the proteome of spelt and bread wheat flour representing targets for research on wheat sensitivities. *Sci. Rep.* **10**, 14677 (2020).
- Zimmermann, J. et al. Comprehensive proteome analysis of bread deciphering the allergenic potential of bread wheat, spelt and rye. *J. Proteom.* **247**, 104318 (2021).
- Afzal, M. et al. Characterization of 150 Wheat Cultivars by LC-MS-Based Label-Free Quantitative Proteomics Unravels Possibilities to Design Wheat Better for Baking Quality and Human Health. *Plants* **10**, 424 (2021).
- Felber, J. et al. Aktualisierte S2k-Leitlinie Zöliakie der Deutschen Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten (DGVS). *Z. für Gastroenterologie* **60**, 790–856 (2022).
- Sergi, C., Villanacci, V. & Carroccio, A. Non-celiac wheat sensitivity: rationality and irrationality of a gluten-free diet in individuals affected with non-celiac disease: a review. *BMC Gastroenterol.* **21**, 1–12 (2021).
- Pinto-Sanchez, M. I. & Verdu, E. F. Non-celiac gluten or wheat sensitivity: It's complicated! *Neurogastroenterol. Motil.: Off. J. Eur. Gastrointest. Motil. Soc.* **30**, e13392 (2018).
- Volta, U. et al. Nonceliac Wheat Sensitivity: An Immune-Mediated Condition with Systemic Manifestations. *Gastroenterol. Clin. North Am.* **48**, 165–182 (2019).
- Aufiero, V. R., Sapone, A. & Mazzarella, G. Diploid Wheats: Are They Less Immunogenic for Non-Celiac Wheat Sensitive Consumers? *Cells* **11**, 2389 (2022).
- Mari, A., Rasi, C., Palazzo, P. & Scala, E. Allergen databases: Current status and perspectives. *Curr. Allergy Asthma Rep.* **9**, 376–383 (2009).
- Longin, C. F. H. et al. Comparative study of hulled (einkorn, emmer, and spelt) and naked wheats (durum and bread wheat): agronomic performance and quality traits. *Crop Sci.* **56**, 302–311 (2016).
- Call, L. et al. Effects of species and breeding on wheat protein composition. *J. Cereal Sci.* **93**, 102974 (2020).
- Geisslitz, S. et al. Wheat ATIs: Characteristics and Role in Human Disease. *Front. Nutr.* **8**, <https://doi.org/10.3389/fnut.2021.667370> (2021).
- Pickert, G. et al. Wheat Consumption Aggravates Colitis in Mice via Amylase Trypsin Inhibitor-mediated Dysbiosis. *Gastroenterology* **159**, <https://doi.org/10.1053/j.gastro.2020.03.064> (2020).
- Sievers, S., Rohrbach, A. & Beyer, K. Wheat-induced food allergy in childhood: ancient grains seem no way out. *Eur. J. Nutr.* **59**, 2693–2707 (2020).
- Larré, C. et al. Assessment of allergenicity of diploid and hexaploid wheat genotypes: Identification of allergens in the albumin/globulin fraction. *J. Proteom.* **74**, 1279–1289 (2011).
- Lombardo, C. et al. Study on the Immunoreactivity of Triticum monococcum (Einkorn) Wheat in Patients with Wheat-Dependent Exercise-Induced Anaphylaxis for the Production of Hypoallergenic Foods. *J. Agric. Food Chem.* **63**, 8299–8306 (2015).
- Alvisi, P. et al. Responses of blood mononucleated cells and clinical outcome of non-celiac gluten sensitive pediatric patients to various cereal sources: a pilot study. *Int. J. Food Sci. Nutr.* **68**, 1005–1012 (2017).
- Ianiro, G. et al. A Durum Wheat Variety-Based Product Is Effective in Reducing Symptoms in Patients with Non-Celiac Gluten Sensitivity: A Double-Blind Randomized Cross-Over Trial. *Nutrients* **11**, <https://doi.org/10.3390/nu11040712> (2019).
- Bordoni, A., Danesi, F., Di Nunzio, M., Taccari, A. & Valli, V. Ancient wheat and health: a legend or the reality? A review on KAMUT khorasan wheat. *Int. J. Food Sci. Nutr.* **68**, 278–286 (2017).
- Shewry, P. R. Do ancient types of wheat have health benefits compared with modern bread wheat. *J. Cereal Sci.* **79**, 469–476 (2018).
- Carroccio, A. et al. Wheat Consumption Leads to Immune Activation and Symptom Worsening in Patients with Familial Mediterranean Fever: A Pilot Randomized Trial. *Nutrients* **12**, <https://doi.org/10.3390/nu12041127> (2020).

59. Picascia, S. et al. In Celiac Disease Patients the In Vivo Challenge with the Diploid Triticum monococcum Elicits a Reduced Immune Response Compared to Hexaploid Wheat. *Mol. Nutr. Food Res.* **64**, 1901032 (2020).
60. Ziegler, J. U. et al. Lutein and Lutein Esters in Whole Grain Flours Made from 75 Genotypes of 5 Triticum Species Grown at Multiple Sites. *J. Agric. Food Chem.* **63**, 5061–5071 (2015).
61. Ziegler, J. U., Schweiggert, R. M., Würschum, T., Longin, C. F. H. & Carle, R. Lipophilic antioxidants in wheat (*Triticum* spp.): a target for breeding new varieties for future functional cereal products. *J. Funct. Foods* **20**, 594–605 (2016).
62. Zeibig, F., Kilian, B. & Frei, M. The grain quality of wheat wild relatives in the evolutionary context. *Theor. Appl. Genet.*, 1–20, <https://doi.org/10.1007/s00122-021-04013-8> (2021).
63. Longin, C. F. H. et al. Mineral and Phytic Acid Content as Well as Phytase Activity in Flours and Breads Made from Different Wheat Species. *Int. J. Mol. Sci.* **24**, 2770 (2023).
64. CIMMYT. 1990-91 CIMMYT World Wheat Facts and Trends: Wheat and Barley Production in Rainfed Marginal Environments of the Developing World (CIMMYT, Mexico, D.F., 1991).
65. Bencze, S. et al. Re-Introduction of Ancient Wheat Cultivars into Organic Agriculture—Emmer and Einkorn Cultivation Experiences under Marginal Conditions. *Sustainability* **12**, 1584 (2020).
66. Manza, L. L., Stamer, S. L., Ham, A.-J. L., Codreanu, S. G. & Liebler, D. C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **5**, 1742–1745 (2005).
67. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
68. Distler, U. et al. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* **11**, 167–170 (2013).
69. Distler, U., Łacki, M. K., Schumann, S., Wanninger, M. & Tenzer, S. Enhancing Sensitivity of Microflow-Based Bottom-Up Proteomics through Postcolumn Solvent Addition. *Anal. Chem.* **91**, 7510–7515 (2019).
70. Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. & Geromanos, S. J. Absolute Quantification of Proteins by LCMSE: A Virtue of Parallel ms Acquisition \*S. *Mol. Cell. Proteom.* **5**, 144–156 (2006).
71. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
72. Cochran, W. G. & Cox, G. *Experimental Designs*. 2nd ed. (Wiley, New York, 1957).
73. Stram, D. O. & Lee, J. W. Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177 (1994).
74. Piepho, H.-P. & Möhring, J. Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* **177**, 1881–1888 (2007).
75. R Core Team. *R: a language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2018).
76. Gilmour, A. R., Gogel, B., Cullis, B. R. & Thompson, R. *ASReml User Guide Release 3.0* (VSN International Ltd, Hemel Hempstead, UK, 2009).
77. Student. The probable error of a mean. *Biometrika* **6**, 1–25 (1908).
78. Levene, H. Robust tests for the equality of variance. In *Contributions to Probability and Statistics*, edited by I. Olkin (Stanford University Press 1960), pp. 278–292.
79. Ward, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
80. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

## ACKNOWLEDGEMENTS

We acknowledge the financial support of the Federal Ministry of Economic Affairs and Energy (BMWi) to M.A. (FKZ: 16KN068825), the German Research Foundation (DFG) to C.F.H.L. (DFG LO 1816–4/1), DFG TE599/3-1 to S.T., DFG Schu 646/17-1 and the Leibniz Foundation (WheatScan, SAW-2016-DFA-2) to D.S.

## AUTHOR CONTRIBUTIONS

M.A. contributed to conceptualization, methodology, analyzed and interpreted data, prepared figures, wrote the first draft; M.S. performed proteomic analyses, analyzed data and contributed to methodology and writing; U.D. contributed to methodology; D.S., S.T., and C.F.H.L. contributed to conceptualization, methodology, writing and funding acquisition. All authors read and approved the final manuscript. M.A., and M.S. contributed equally to this work.

## FUNDING

Open Access funding enabled and organized by Projekt DEAL.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41538-023-00188-0>.

**Correspondence** and requests for materials should be addressed to C. Friedrich H. Longin.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023