



Genetic architecture underlying the expression of eight α -amylase trypsin inhibitors

Khaoula EL Hassouni¹ · Malte Sielaff² · Valentina Curella³ · Manjusha Neerukonda³ · Willmar Leiser¹ · Tobias Würschum⁴ · Detlef Schuppan^{3,5} · Stefan Tenzer² · C. Friedrich H. Longin¹

Received: 19 April 2021 / Accepted: 1 July 2021 / Published online: 10 July 2021
© The Author(s) 2021

Abstract

Key message Wheat cultivars largely differ in the content and composition of ATI proteins, but heritability was quite low for six out of eight ATIs. The genetic architecture of ATI proteins is built up of few major and numerous small effect QTL.

Abstract Amylase trypsin inhibitors (ATIs) are important allergens in baker's asthma and suspected triggers of non-celiac wheat sensitivity (NCWS) inducing intestinal and extra-intestinal inflammation. As studies on the expression and genetic architecture of ATI proteins in wheat are lacking, we evaluated 149 European old and modern bread wheat cultivars grown at three different field locations for their content of eight ATI proteins. Large differences in the content and composition of ATIs in the different cultivars were identified ranging from 3.76 pmol for ATI CM2 to 80.4 pmol for ATI 0.19, with up to 2.5-fold variation in CM-type and up to sixfold variation in mono/dimeric ATIs. Generally, heritability estimates were low except for ATI 0.28 and ATI CM2. ATI protein content showed a low correlation with quality traits commonly analyzed in wheat breeding. Similarly, no trends were found regarding ATI content in wheat cultivars originating from numerous countries and decades of breeding history. Genome-wide association mapping revealed a complex genetic architecture built of many small, few medium and two major quantitative trait loci (QTL). The major QTL were located on chromosomes 3B for ATI 0.19-like and 6B for ATI 0.28, explaining 70.6 and 68.7% of the genotypic variance, respectively. Within close physical proximity to the medium and major QTL, we identified eight potential candidate genes on the wheat reference genome encoding structurally related lipid transfer proteins. Consequently, selection and breeding of wheat cultivars with low ATI protein amounts appear difficult requiring other strategies to reduce ATI content in wheat products.

Communicated by Hermann Buerstmayr.

✉ C. Friedrich H. Longin
friedrich.longin@uni-hohenheim.de

¹ State Plant Breeding Institute, University of Hohenheim, Fruwirthstr. 21, 70599 Stuttgart, Germany

² Institute for Immunology, University Medical Center of the Johannes Gutenberg-University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany

³ Institute of Translational Immunology and Research Center for Immune Therapy, University Medical Center of the Johannes Gutenberg-University Mainz, Langenbeckstr. 1, 55131 Mainz, Germany

⁴ Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Fruwirthstr. 21, 70599 Stuttgart, Germany

⁵ Division of Gastroenterology, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Ave, Boston, MA 02215, USA

Introduction

Wheat (*Triticum aestivum* subsp. *aestivum*) is one of the predominant crops globally and the most consumed cereal worldwide (FAO 2019) providing a high nutritional value with fibers, minerals, vitamins, proteins and starch (Shewry 2009). However, wheat consumption can cause adverse health reactions. Besides celiac disease and classical wheat allergy like baker's asthma, a small proportion of wheat consuming subjects appears to suffer from non-celiac wheat sensitivity (NCWS; Schuppan et al. 2015; Catassi et al. 2017). NCWS can be defined as an adverse immune-mediated reaction to wheat products. Patients report intestinal or extra-intestinal symptoms, usually delayed by several hours after consumption of wheat. Wheat amylase trypsin inhibitors (ATIs) play a prominent role in the reactions. They are largely resistant to digestion by gastrointestinal proteases and activate intestinal innate immune cells, mainly

macrophages and dendritic cells via the toll like receptor 4 (TLR4; Junker et al. 2012; Cuccioloni et al. 2017; Zevallos et al. 2017). In mouse models of disease, this causes an exacerbation of inflammatory bowel disease (Zevallos et al. 2017; Pickert et al. 2020), fatty liver disease (Ashfaq-Khan et al. 2019), classical food or respiratory allergies (Bellinghausen et al. 2019; Zevallos et al. 2019), and even Alzheimer's disease (Dos Santos Guilherme et al. 2020). ATIs are major triggers of classical, immediate-type, immunoglobulin E (IgE)-inducing respiratory and food allergies (Salcedo et al. 2011; Schuppan et al. 2015; Kalunke et al. 2020), and importantly, are now implicated as allergens in a novel, highly prevalent type of IgE-negative food allergy, as a major cause of the irritable bowel syndrome that affects up to 15% of most societies (Fritscher-Ravens et al. 2014, 2019). This IgE-negative, likely eosinophil and T helper 2 cell-mediated intestinal wheat allergy represents a major fraction of patients with NCWS (Carroccio et al. 2013).

Wheat ATIs consist of three subfamilies: tetramers, dimers, and monomers (Silano et al. 1973, 1977; Oda et al. 1997; Altenbach et al. 2011) that are all included in the non-gluten (albumin) protein fraction. Over the past years, the major research emphasis regarding wheat protein families associated with digestive health has been on gluten (Caminero and Verdu 2019). This major wheat protein fraction is important for bread and pasta quality but contains also the epitopes for celiac disease with numerous studies investigating its distribution in different wheat species and cultivars (Call et al. 2020; Geisslitz et al. 2020; Pronin et al. 2020a). Roughly summarized, gluten content decreased slightly comparing old with modern wheat cultivars, and gluten composition changed toward more glutenins and less gliadins mainly due to the favorable processing properties of glutenins (Call et al. 2020; Pronin et al. 2020b). In contrast to gluten, ATIs have only recently come into the focus of clinical wheat research. Studies published prior to 2012 mainly related to characterization of the structure and the function of ATIs (Oda et al. 1997; Franco et al. 2002), and their important if not prominent role in immediate type, IgE-mediated respiratory wheat allergy (baker's asthma). The discovery that ATIs activate immune cells via TLR4 in vitro and in vivo has greatly stimulated research related to their biochemistry, biology and physiological effects (Salcedo et al. 2011; Junker et al. 2012; Carroccio et al. 2013; Fritscher-Ravens et al. 2014, 2019; Fasano et al. 2015; Schuppan et al. 2015; Schuppan and Zevallos 2015; Catassi et al. 2017; Cuccioloni et al. 2017; Zevallos et al. 2017, 2019; Dinu et al. 2018; Reig-Otero et al. 2018; Tundo et al. 2018; Ashfaq-Khan et al. 2019; Bellinghausen et al. 2019; Schuppan and Gisbert-Schuppan 2019; Kalunke et al. 2020; Pickert et al. 2020). Moreover, ATI quantification in various wheats was performed using different extraction protocols (Prandi et al.

2013; Zevallos et al. 2017; Geisslitz et al. 2018; Bose et al. 2019). The currently largest study by Geisslitz et al. (2020) compared the content of 13 ATI proteins investigating eight cultivars among five wheat species, including hexaploid spelt (*T. aestivum ssp. spelta*), tetraploid durum (*T. turgidum ssp. durum*) and emmer (*T. turgidum ssp. dicoccum*), and diploid einkorn (*T. monococcum ssp. monococcum*) grown at three different environments, with very low concentrations of the measured ATI proteins in einkorn, while the other species containing a comparable level of total ATIs. However, significant variations for the different ATIs within each wheat species, with moderate heritabilities, were observed. Nevertheless, this study was based on only eight cultivars per species and estimates of genetic variances and heritabilities have normally large errors or confidence intervals. Moreover, to the best of our knowledge, environmental effects on ATI compositions have not yet been investigated in bread wheat, and studies on the genetic architecture of ATI proteins in wheat are lacking.

We therefore investigated a highly diverse set of 149 wheat cultivars tested at three different field locations for their amount of eight ATI proteins. Additionally, a genome-wide association mapping approach was performed using 22,220 Diversity Arrays Technology (DArT) markers. Our objectives were to (1) examine the genotype and the environment effect on the expression of eight major ATIs, (2) evaluate the correlation of these ATI proteins with important quality parameters, and (3) elaborate the genetic architecture underlying the expression of these eight ATIs.

Materials and methods

Plant material and field experiments

A total of 149 bread wheat ($2n=6x=42$, AABBDD) cultivars were used in this study, which were registered between 1921 and 2013 and originate from different European countries. The list of the cultivars and their details are provided in Table S1. Field trials were conducted in one winter cropping season (2015–2016) at three locations, Hohenheim (HOH, 48°43'07.3"N 9°11'08.7"E, altitude 403 m), Oberer Lindenhof (OLI, 48°28'19.0"N 9°18'29.3"E, altitude 700 m) and Eckartsweier (EWE, 48°32'52.4"N 7°52'32.5"E, altitude 140 m) in Germany. Trials were arranged in partially replicated (P-rep) designs with a replication factor of 1.125 (Williams et al. 2011) and a net plot size of 1.25 m². Plants were sown in October and harvested in July. During the growing season, standard cultivation practices were adopted of intensive wheat production applying 200 kg/ha of nitrogen fertilizer including Nmin, growth regulators and fungicides.

Protein extraction

The seeds of all the wheat samples were cleaned using the Mini-Petkus seed cleaner and then milled using a laboratory mill equipped with a 1 mm sieve (Cyclotec 1093, FOSS, Hillerød, Denmark) to obtain the whole-grain flour. ATIs were quantitatively extracted from 1 g of whole-grain flour using 5 mL of extraction buffer (10 mM sodium bicarbonate, 500 mM sodium chloride, pH 7.8) with constant spinning at 4 °C overnight. The suspension was centrifuged at 4fSubsequently, the best linear unbiased $600\times g$ for 30 min, the supernatant collected and the procedure was repeated with an additional 5 mL of extraction buffer. Supernatants were combined and sterile-filtered (0.22 μm). Our prior study showed that this procedure extracted >90% of ATIs, while maintaining their native protein conformation and solution stability (Sielaff et al. 2021).

ATI quantification by liquid chromatography-mass spectrometry

Proteomic sample preparation, liquid chromatography-mass spectrometry (LC–MS) analysis, data processing and absolute quantification of ATIs using isotope-labeled standard peptides were performed as detailed before (Sielaff et al. 2021).

5 μL of flour extract was spiked with 30 pmol of $^{13}\text{C}_6$ -arginine- and $^{13}\text{C}_6$ -lysine-labeled ATI QconCAT protein, an artificial concatemer of standard peptides (Beynon et al. 2005) used for ATI quantification (Addgene plasmid # 163,957; <http://n2t.net/addgene:163957>; RRID: Addgene_163957). After denaturation using 0.1% (w/v) RapiGest (Waters Corporation, Milford, MA, USA), proteins were reduced, alkylated and digested overnight using MS grade trypsin (Promega, Madison, WI, USA) at a protease-to-protein ratio of 1:25 (w/w). RapiGest was hydrolyzed by addition of trifluoroacetic acid and precipitates were removed by centrifugation. Peptides were further desalted using Sep-Pak tC_{18} cartridges (Waters Corporation) and lyophilized. Prior to LC–MS analysis, peptides were dissolved in 20 μL of 0.1% (v/v) formic acid.

Furthermore, a reference mix of extracts was generated and used to titrate a dilution series from 40 to 0.625 μL . Each sample of the dilution series was spiked with 30 pmol of QconCAT protein and processed as described before. The reference samples were later used to generate external calibration curves for peptide quantification.

LC–MS analyses of tryptic peptides were performed using a nanoACQUITY UPLC system (Waters Corporation) connected to a SYNAPT G2-S mass spectrometer (Waters Corporation) via a NanoLockSpray dual electrospray ionization source (Waters Corporation). 0.5 μL of peptides were directly injected onto a HSS-T3 300 $\mu\text{m}\times 100\text{ mm}$, 1.8 μm

reversed-phase column (Waters Corporation) and eluted using a gradient mobile phase at a flow rate of 8 $\mu\text{L}/\text{min}$ for 15 min. Mobile phase solvent A was water with 0.1% (v/v) formic acid. Solvent B was acetonitrile with 0.1% (v/v) formic acid, which was gradually increased from 1 to 36% (v/v) during elution. The column temperature was 55 °C. Post-column addition of 25% (v/v) dimethyl sulfoxide in acetonitrile at a flow rate of 1 $\mu\text{L}/\text{min}$ was performed as described before (Distler et al. 2019). 250 fmol/ μL [Glu1]-Fibrinopeptide B was directly infused into the electrospray ionization source via the reference sprayer using a flow rate of 1.5 $\mu\text{L}/\text{min}$.

Mass spectra were acquired by alternating between low (MS) and elevated energy scans (MS^E). Acquisition time was 0.4 s in each scan with an interscan delay of 0.05 s. During MS scans, a constant collision energy of 4 eV was applied, while the collision energy was ramped from 16 to 40 eV during MS^E scans. The doubly charged monoisotopic ion of [Glu1]-Fibrinopeptide B was used as lock mass by sampling the reference sprayer in 30 s-intervals.

Raw data of standard samples were processed using ProteinLynx Global Server v3.0.2 (PLGS, Waters Corporation) and searched against a database containing *T. aestivum* proteins (UniProtKB release 2019_11, taxon ID: 4546, 142,969 entries + potential contaminants), specifying trypsin as protease, allowing two missed cleavages per peptide and defining carbamidomethylation as fixed and methionine oxidation as variable modification. In addition, isotope-labeled lysine and arginine were allowed as variable modification. The false discovery rate (FDR) was estimated by searching a reversed protein database and an FDR cut-off of 0.01 was applied.

The results were used to build a spectral library for the QconCAT peptides and unlabelled forms in Skyline v20.1.0.155. Afterward, targeted extraction of raw data, peak area integration, generation of external standard curves and absolute quantification at the precursor ion level were performed using Skyline as described in Sielaff et al. (2021).

The quantities of following ATI types were reported as average of indicated peptide quantities: monomeric ATI 0.28 (SVYQELGVR, LQCVGSQVPEAVLR), dimeric ATI 0.19 (DCCQQLAHISEWCR, LTAASITAVCR), dimeric ATI 0.19-like (VPALPGCRPVLK, LTAASITAVCK), tetrameric ATI CM1 (EYVAQQTTCGISISGSVSTEPGNTPR), ATI CM2 (EYVAQQTTCGVIVGSPVSTEPGNTPR), ATI CM3 (SGNVGESGLIDLPGCPR, YFIALPVPSQPVDPR), ATI CM16 (QQCCGELANIPQQCR, IETPGSPYLAK) and ATI CM17 (QECCEQLANIPQQCR, IEMPGPPYLAK). Units of measure are expressed as pmol per 5 μL of ATI extract used for proteomic analysis, equivalent to 0.5 mg of extracted wheat flour.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE

(Perez-Riverol et al. 2019) partner repository with the dataset identifier PXD023654.

Quality traits measurements

The analysis of the quality traits asparagine content (ASP), protein content (PC), sedimentation volume (SDS), sulfur content (SC), thousand kernel weight (TKW), test weight (TW) and falling number (FN) were described in detail in a previous study (Rapp et al. 2018).

Statistical analysis

The phenotypic data were analyzed according to the following linear mixed model:

$$y_{ijkl} = \mu + g_i + env_j + g_i : env_j + r_{jk} + b_{jkl} + e_{ijkl}$$

where y_{ijkl} was the phenotypic observation of the i th genotype tested at the j th environment in the k th replication and in l th block, μ was the general mean, g_i was the genotypic effect of the i th genotype, env_j was the effect of the j th environment, $g_i : env_j$ was the genotype-by-environment interaction, r_{jk} was the effect of the k th replication at the j th environment, b_{jkl} was the effect of the l th block at the k th replication at the j th environment and e_{ijkl} was the residual error. Genotype, genotype by environment interaction and block effects were considered as random, and environment and replication as fixed effects model terms. The significance of random terms was tested by model comparison using a likelihood ratio test, and Wald test was used for significance of the fixed terms.

Subsequently, the best linear unbiased estimates (BLUEs) were estimated across the three environments (locations) assuming fixed genetic effects. The least significant difference (LSD) was determined for each trait at a significance level of 0.05. Broad sense heritability (H^2) was calculated following Piepho and Möhring (2007) as: $H^2 = 1 - \frac{\vartheta}{2\sigma_G^2}$, where ϑ is the mean variance of a difference of two best linear unbiased predictors and σ_G^2 the genetic variance. Pearson correlation coefficients were estimated for all traits using the BLUEs across environments and the R package ‘corrplot’ (Wei and Simko 2017).

Box-and-whisker plots were constructed using the R package ggplot2 to investigate the distribution of the cultivar panel based on the year of registration and origin of the wheat cultivars. To test whether the ATI content has changed between the groups of registration periods and varietal origin, we compared the groups using non-parametric multiple comparison test Kruskal-Wallis with the R package ‘pgrimess’. All phenotypic analyses were conducted using the statistical software R (R Development Core Team 2016) and software package ASReml-R 3.0 (Gilmour et al. 2009).

Genomic data and genotyping

Details of the genotyping approach are described in Rapp et al. (2018). Briefly, the cultivars were genotyped by genotyping-by-sequencing at Diversity Arrays Technology (Yarralumla, Australia) using the Wheat GBS 1.0 assay (DARtseq). DARtseq delivers two types of data, codominant SNP (S) and dominant DARt (D) markers. SNP and DARt markers showing more than 25% missing data and having minor allele frequencies lower than 5% were removed. The remaining missing markers were imputed using the package LinkImpute (Money et al. 2015). A total of 41,604 high-quality markers were obtained with only 22,122 markers having known genetic map positions (Li et al. 2015). Therefore, a chromosomal position was assigned to the significantly associated unmapped markers (98) based on the highest ($r^2 > 0.4$) linkage disequilibrium (LD) with the mapped markers. Finally, we ended up with 22,220 markers with a map position. The physical positions of the markers were determined by BLASTing the marker sequences against the wheat genome IWGSC RefSeq v1.0 (The International Wheat Genome Sequencing Consortium 2018).

Genome-wide association study (GWAS) and genomic prediction

GWAS was carried out using the R package GenABEL, (Aulchenko et al. 2007) where a mixed linear model was fitted incorporating marker data and a kinship matrix. Principal coordinates analysis did not show any clusters; therefore, principal coordinates were not included in the model. We used an explorative significance threshold of $P < 0.001$ and a Bonferroni-corrected threshold of $P < 0.05$ to identify significant marker-trait associations.

To obtain the genotypic variance explained by each quantitative trait locus (QTL) and across all QTL, we used a linear model fitting the significant QTL ordered based on the strength of their association. The explained genotypic variance (p_G) was calculated as: $p_G = \frac{R^2_{adj}}{H^2}$ (Utz et al. 2000; Würschum et al. 2016), where R^2_{adj} is the adjusted R^2 from the linear model and H^2 is the heritability of the trait. Only the most strongly associated markers, which explained more than 1% of the genotypic variance, were declared as putative QTL and reported in the manuscript.

Genomic prediction was conducted based on Ridge Regression best linear unbiased prediction (RR-BLUP) method with fivefold cross-validation using the R package ‘rrBLUP’ (Endelman 2011). In addition, weighted Ridge Regression-BLUP was performed by including the identified QTL explaining more than 10% of the genetic variance as fixed effects in the model (Zhao et al. 2014).

For candidate gene search, the latest publicly available wheat genome (IWGSC RefSeq v1.0) and gene functional

annotation information were downloaded from the URGI database (Alaux et al. 2018; available at: <https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations>). High (HC) and low confidence (LC) genes were extracted from the identified chromosomal regions of the significant QTL explaining more than 10% of genotypic variance. The genes with the functional annotation similar to the different domains of ATIs were selected as potential candidate genes.

Results

Phenotypic characterization of ATI proteins in wheat

We detected a wide range of genotypic values for the content of all ATIs across the 149 wheat cultivars (Table 1). For instance, content of ATI 0.28 ranged from 4.96 pmol for cultivar 'Mv Zelma' to 31.77 pmol for cultivar 'Potenzial'. Similarly, a wide variation was determined for the content of ATI 0.19 ranging from 14.67 pmol for cultivar 'Expert' to 80.40 for cultivar 'Granada'. For the content of the other six investigated ATI proteins, differences were less pronounced, with a roughly twofold difference between minimum and maximum content. Summing up all eight ATI proteins, total ATI showed an up to twofold difference across the cultivars ranging from 98.55 pmol for cultivar 'Cezanne' up to 195.12 pmol for cultivar 'Slejpner' (Fig. 1a). The average contribution of single ATIs to average total ATI content across the 149 wheat cultivars was 31% ATI 0.19, 13% ATI 0.28, 12% ATI CM16, 11% ATI 0.19-like, 11% ATI CM3, 10% ATI CM17, 8% ATI CM1 and 4% ATI CM2. However, this percentage contribution of individual ATI proteins also largely varied across the 149 cultivars (Fig. 1b). For

instance, (dimeric) ATI 0.19 contributed between 14 and 46% and tetrameric ATI CM16 contributed between 11.84 and 24.81% of the total ATI content among the 149 cultivars.

These large differences led to highly significant genotypic variances for all eight ATIs and total ATI content (Table 1). For ATI 0.28, ATI 0.19 and ATI 0.19-like, very low variances due to genotype-by-environment interaction were determined. For the other ATI proteins and total ATI content, a considerable genotype-by-environment interaction was found. Furthermore, variances due to residual error were as high as genotypic variances or even higher for almost all ATIs. This led to low to medium estimates of broad sense heritability ranging from 0.36 for total ATI content to 0.79 for ATI 0.28 across all locations. At single locations, the heritabilities (repeatabilities) ranged from 0.14 for ATI 0.19-like in EWE to 0.81 for CM1 in OLI (data not shown). Except for all CM type ATIs, correlation coefficients across different ATI proteins were low (Table 2). Similarly, all ATIs showed only weak correlations with important quality traits commonly assessed in wheat breeding (Table 2).

The investigated wheat cultivars were registered from 1921 to 2013 and originate from different European countries. We split them in different groups for time periods between 1921 and 1960 and from 1961 onward in decades leading to group sizes between 13 and 23 except for the last decade, where 52 varieties were assessed (Fig. 2). We found a large variability for individual ATIs and for total ATI content in each time period, and no temporal trend across the decades. Similarly, we formed groups of origin across seven regions in Europe (Fig. S1), but group sizes largely varied with an imbalance due to a large number of German cultivars. In each group, we also found a wide variation in ATI contents but no regional trend. According to the

Table 1 Summary statistics for the eight ATI proteins (monomeric ATI 0.28, dimeric ATI 0.19 and ATI 0.19-like, tetrameric ATI CM1, ATI CM2, ATI CM3, ATI CM16, ATI CM17) measured among 149 bread wheat cultivars tested in three environments

Trait	ATI 0.28 (pmol)	ATI 0.19 (pmol)	ATI 0.19-like (pmol)	ATI CM1 (pmol)	ATI CM2 (pmol)	ATI CM 3 (pmol)	ATI CM 16 (pmol)	ATI CM 17 (pmol)	Total ATI (pmol)
Min	4.96	14.67	9.42	7.20	3.76	11.04	11.84	9.75	98.55
Mean	19.49	43.97	15.38	11.09	5.64	15.70	16.89	14.56	143.04
Max	31.77	80.40	21.96	16.82	9.64	22.12	24.81	22.44	195.12
LSD	5.67	26.36	4.47	2.64	1.46	4.70	5.12	3.92	43.40
σ_G^2	15.27***	49.27***	3.03***	1.04***	0.63***	2.10***	2.24***	2.21***	127.87**
$\sigma_{G \times E}^2$	0.09	0.00	0.00	0.58	0.27*	1.96	3.32*	1.39	87.30
σ_e^2	12.22	236.08	6.97	1.87	0.46	5.83	5.47	3.94	566.65
H^2	0.79	0.37	0.55	0.55	0.70	0.44	0.42	0.54	0.36
E effect [†]	***	***	*	***	***	***	ns	**	***

LSD least significant difference at 5% probability level, σ_G^2 genotypic variance, $\sigma_{G \times E}^2$ genotype-by-environment interaction variance, σ_e^2 error variance, H^2 heritability

*, **, ***Significant at the 0.05, 0.01 and 0.001 probability levels, respectively

[†]Significance of the environment (E) effect according to Wald test, which was taken as fixed term in the statistical model; ns = non significant

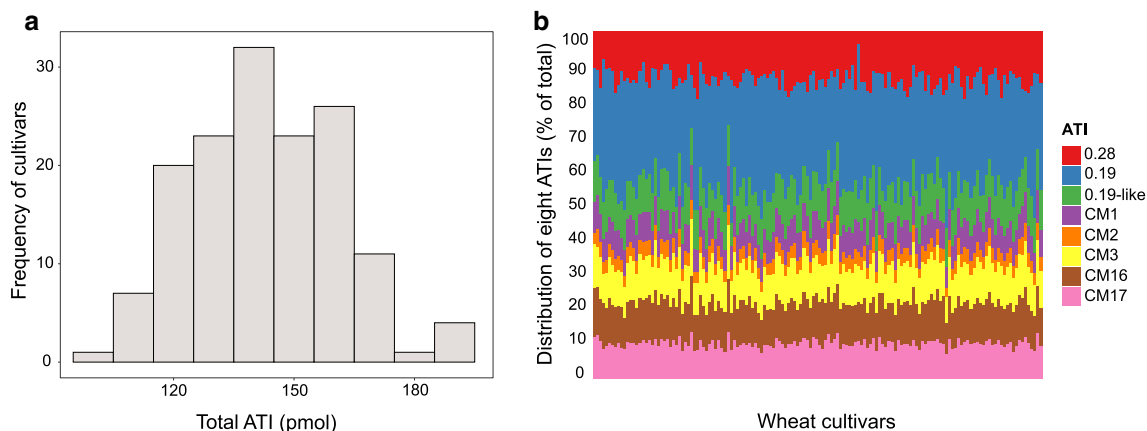


Fig. 1 Variability of total ATI content, which is the sum of all eight measured individual ATI proteins **a** and the variability of ATI composition per cultivar by setting total ATI content to 100% **b** across the 149 investigated wheat cultivars

Table 2 Pearson's correlation coefficients among eight measured ATI proteins as well as between ATI proteins, total ATI content and quality traits assessed using the same set of bread wheat cultivars tested in three environments

	ATI 0.28	ATI 0.19	ATI 0.19-like	ATI CM1	ATI CM2	ATI CM3	ATI CM16	ATI CM17	Total ATI
ATI 0.28									
ATI 0.19	0.16								
ATI 0.19-like	0.28***	0.42***							
ATI CM1	0.18*	0.27***	0.41***						
ATI CM2	0.10	0.03	0.28***	0.85***					
ATI CM3	0.28***	0.15	0.3***	0.83***	0.78***				
ATI CM16	0.21**	0.37***	0.35***	0.78***	0.69***	0.83***			
ATI CM17	0.18*	0.14	0.31***	0.87***	0.87***	0.87***	0.77***		
Total ATI	0.47***	0.81***	0.62***	0.68***	0.47***	0.62***	0.72***	0.60***	
ASP	-0.05	0.09	-0.19*	-0.14	-0.07	-0.16	-0.09	-0.10	-0.03
PC	0.10	0.19*	0.04	0.17*	0.20*	0.11	0.14	0.20	0.23*
SDS	-0.02	0.05	-0.03	-0.06	-0.12	-0.08	-0.02	-0.12	-0.02
SC	0.10	0.18*	-0.05	0.39***	0.40***	0.41***	0.39***	0.39***	0.32***
TKW	-0.09	-0.21**	-0.09	0.01	0.09	0.19*	0.07	0.07	-0.12
TW	0.04	0.01	-0.14	-0.01	-0.11	0.05	0.12	-0.03	0.01
FN	0.09	-0.07	-0.03	-0.19*	-0.09	-0.20	-0.13	-0.11	-0.10

ASP asparagine content; PC protein content; SDS sedimentation volume; SC sulfur content; TKW thousand kernel weight; TW test weight; FN falling number

*, **, ***Significant at the 0.05, 0.01 and 0.001 probability levels, respectively

Kruskall-Wallis test, no significant difference (for $p < 0.05$) between the grouping either for registration years or origin was found.

The genetic architecture underlying ATI protein content

A genome-wide association mapping using 22,220 polymorphic markers delivered a total of 68 QTL significant at $p < 0.001$ across the eight ATIs and total ATI content

(Table S2, Fig. S2). Among them, 30 QTL explained more than 10% of the genotypic variation for ATI 0.28 on chromosome 6B, for ATI 0.19 on chromosomes 1B and 3B, for ATI 0.19-like on chromosome 3B, for ATI CM1 on chromosomes 1A, 1B and 4B, for ATI CM2 on chromosomes 1B, 7B and 1D, for ATI CM3 on chromosomes 1A, 5A, 4B and 6D, for ATI CM16 on chromosomes 5A, 6B and 7B, for ATI CM17 on chromosomes 1B, 4B and 7D as well as for total ATI on chromosomes 1A, 6A, 1B, 2B, 3B and 2D. Furthermore, QTL in two chromosomal regions surpassed the Bonferroni

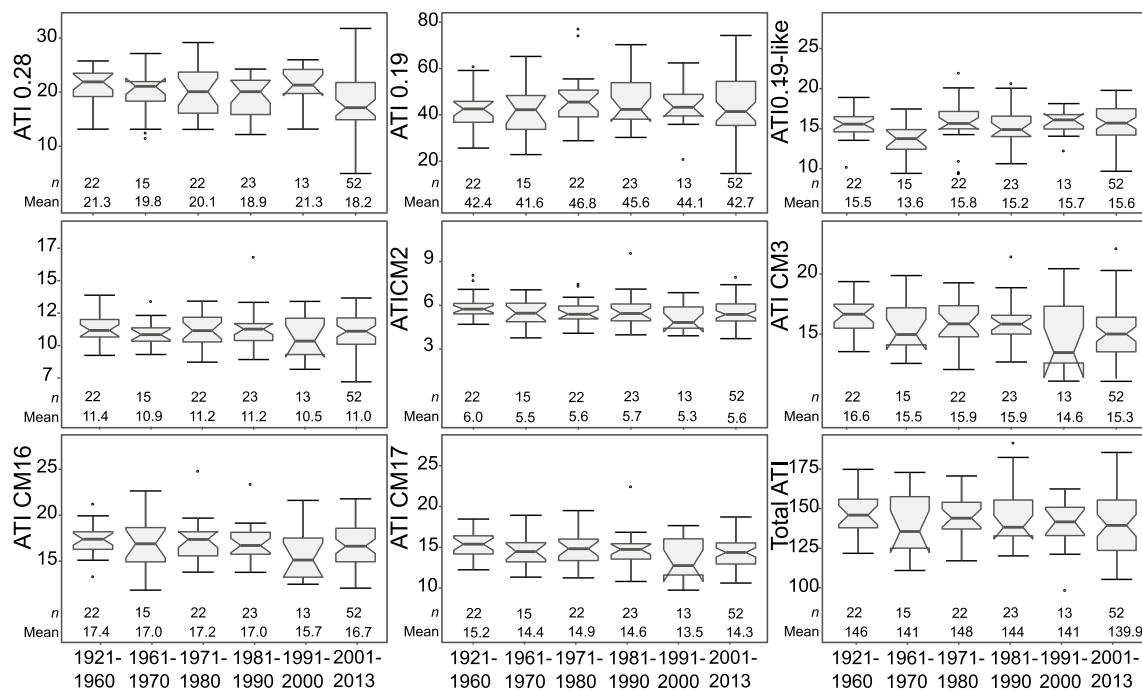


Fig. 2 Boxplots showing the amount of different ATI proteins in the wheat cultivars depending on the period, when the respective wheat cultivars were officially registered. Numbers in the plot indicate the

mean values and the number of cultivars in each group. Two cultivars with lacking registration information were omitted

significance threshold of $p < 0.05$ ($-\log_{10} p$ value = 5.65; Table 3, Fig. 3). These two major QTL explained 68 and 71% of the genotypic variation for ATI 0.28 and ATI 0.19-like, respectively.

After this initial genome scan, we investigated in more detail the chromosomal regions harboring the 30 significant QTL, which explained more than 10% of the genotypic variation (Fig. 4, Fig. S3). For these regions, we extracted high (HC) and low confidence (LC) genes from the bread wheat reference genome (IWGSC RefSeq v1.0) and evaluated these as potential candidate genes with functional annotations similar to the different domains of ATIs. With that approach, eight potential candidate genes were identified out of a total of 1,081 genes harbored in the target regions (Table 4, Table S3). These candidate genes were TraesCS3B02G111100 and TraesCS3B02G111200 for ATI 0.19-like located between 75 and 80 Mbp on chromosome 3B, TraesCS1A02G048700 for

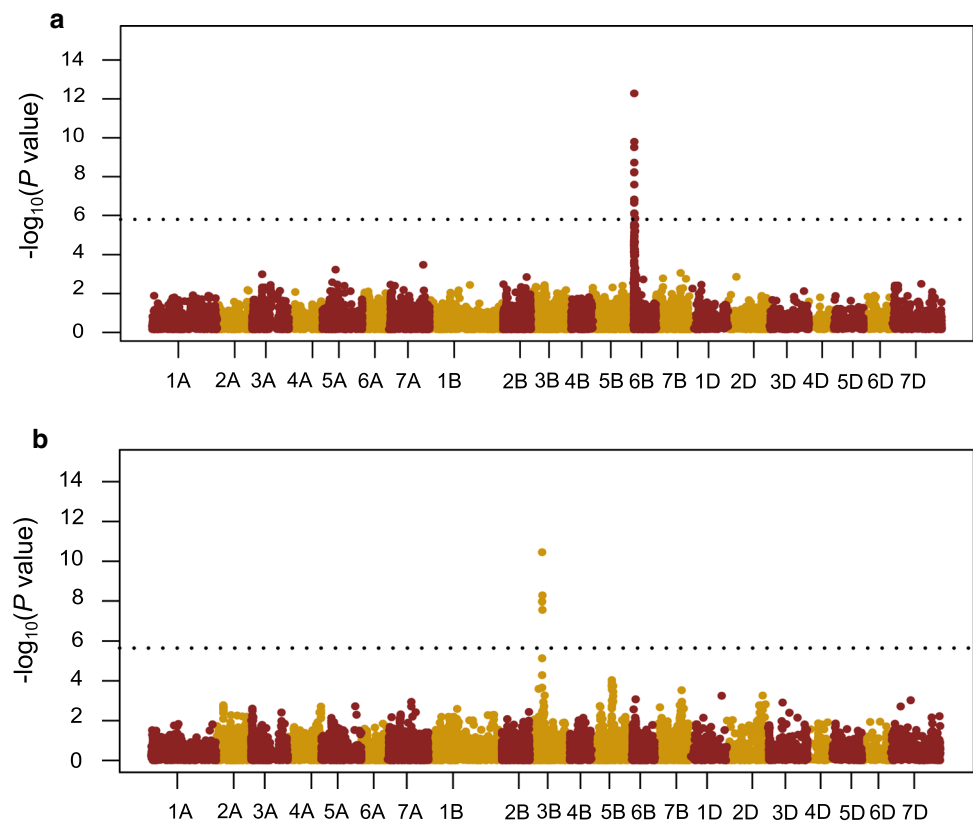
ATI CM1 located in the region 29.5–30.5 Mbp of chromosome 1A, TraesCS7B02G072000 located in the region between 72 and 85 Mbp on chromosome 7B, TraesCS1D02G163900 and TraesCS1D02G164000 located on chromosome 1D between 234 and 235.5 Mbp for ATI CM2, as well as TraesCS3B02G170800 located on chromosome 3B between 171 and 174 Mbp for total ATI content. No potential candidate genes with annotations in the Pfam and InterPro databases with relation to ATI could be identified for the QTL detected for the ATI proteins 0.28, 0.19, CM3, CM16 and CM17. Finally, we investigated the linkage disequilibrium (LD) pattern in the chromosomal region of the two major QTL for ATI 0.28 and ATI 0.19-like (Fig. 4, Fig. S3). The region of the major QTL detected for ATI 0.19-like on chromosome 3B harbors many genes (63) including the two candidate genes TraesCS3B02G111100 and TraesCS3B02G111200 (Fig. 4). In contrast to our expectation of a LD pattern with high LD

Table 3 Marker-trait associations surpassing Bonferroni threshold (5.65) indicating putative quantitative trait loci (QTL) for ATI 0.28 and ATI 0.19-like

Trait	Marker	Chr	Gen Pos (cM)	Phy Pos (bp)	LOD	P_G	α -Effect
ATI 0.28	1106155D	6B	6.45	2,094,331	13.71	68.73	3.30
ATI 0.19-like	2275974D	3B	93.78	78,749,798	10.45	70.55	-2.67
	1123753S	3B	93.78	78,136,767	7.97	7.58	-1.91

Chr. Chromosome, Gen Pos chromosome position in cM, Phy Pos sequence start position in base pairs according to the bread wheat reference genome (IWGSC RefSeq v1.0), P_G proportion of genotypic variance explained by the QTL in percent, and allele substitution (α) effect

Fig. 3 Manhattan plots showing significant marker-trait associations for ATI 0.28 (a) and ATI 0.19-like (b) at Bonferroni-corrected significance threshold of $P < 0.05$. The x-axis shows the DArTseq markers on 21 chromosomes based on the genetic map positions (cM) and the y-axis shows the P values on a $-\log_{10}$ scale



between our significant markers and the markers being closest to the two candidate genes in that small chromosomal region, we determined a break of LD pattern within this region. The markers closest to the candidate genes were only in very weak LD to the two highly significant markers. For the major QTL on chromosome 6B for ATI 0.28, we could neither find any potential candidate genes with an annotation similar to an ATI nor a strong LD pattern in that chromosomal region (Fig. S3).

Finally, we performed a genome-wide prediction approach for all eight ATIs and total ATI content (Fig. S4). For all ATIs, the cross-validated prediction ability was already high using marker-assisted selection based only on the QTL explaining more than 10% of the genotypic variance. Combining these markers with RR-BLUP slightly improved the cross-validated prediction abilities. The highest average prediction ability was achieved with $r = 0.76$ for ATI 0.28 and the lowest with $r = 0.43$ for ATI 0.19.

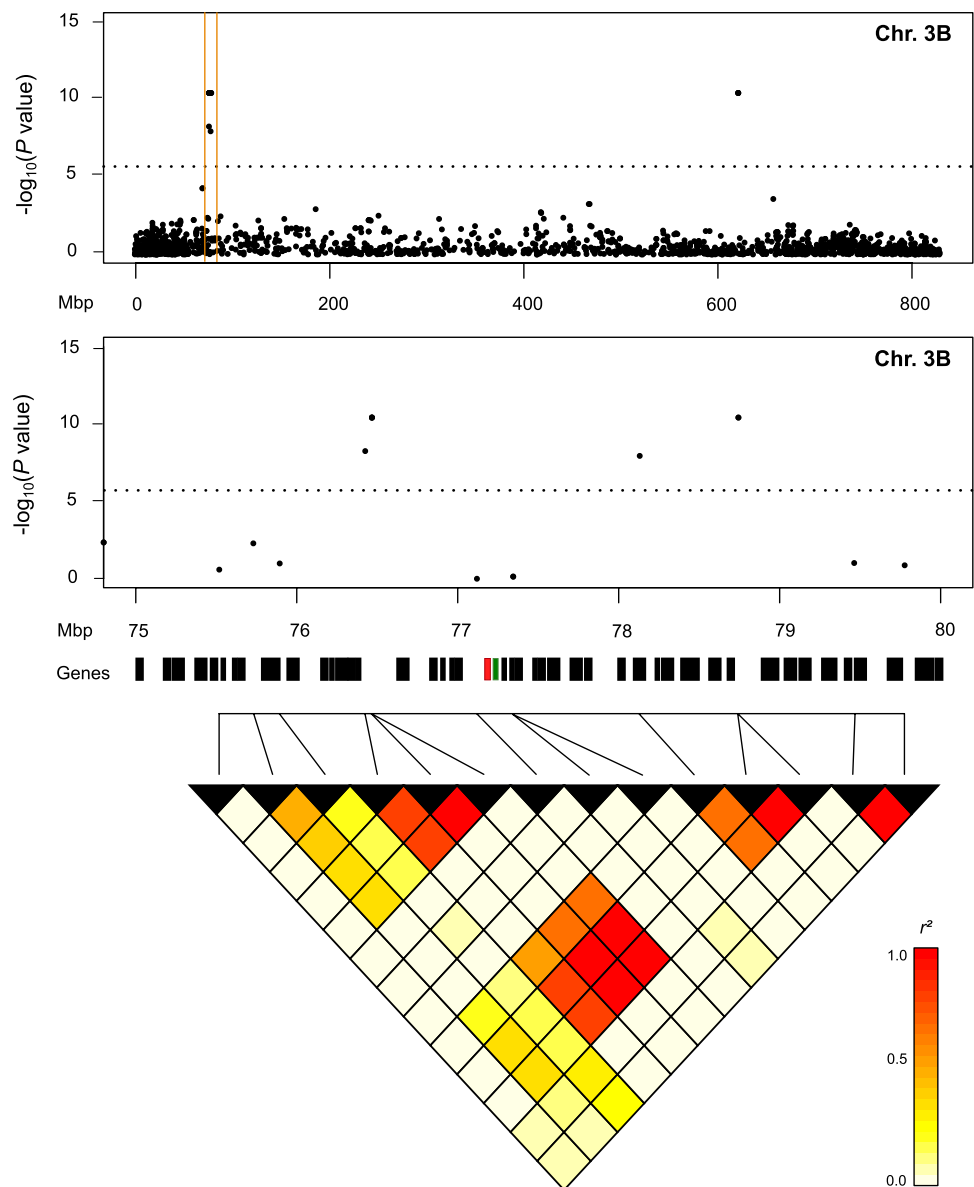
Discussion

Wheat cultivars largely differ in their content and composition of ATIs

The investigated 149 wheat cultivars largely differed in their contents of the eight evaluated (major) ATIs as well

as their total ATI content (Table 1, Fig. 1a). For instance, the content of ATI 0.28 and ATI 0.19 varied more than fivefold between the different cultivars, while for the other ATIs and total ATI content, this variation was up to two- or threefold. Furthermore, contents of the different ATIs were only partly correlated across the cultivars. For all measured tetrameric CM type ATIs, high coefficients of correlation were found (Table 2). Thus, a wheat cultivar low in one CM type ATI trended to be also low in the other CM type ATIs. By contrast, coefficients of correlation between ATI 0.19, ATI 0.19-like and ATI 0.28 as well as between them all and all CM type ATIs were quite low. For instance, the lowest content of ATI 0.28 was identified for the Hungarian cultivar 'Mv Zelma' but with a relatively high content of ATI 0.19. These differences led to a large variation in ATI composition in the different wheat cultivars (Fig. 1b). While few cultivars were quite low across most ATIs leading also to a low total ATI content (e.g., 'Cezanne', 'Akteur', 'Hermann'), others were low in one ATI but high in other ATIs (e.g., 'Mv Zelma', 'Skater'). These results are in agreement with findings of Bose et al. (2020), who compared 23 hexaploid wheat lines for 18 ATIs, Geisslitz et al. (2020) and Call et al. (2020), who quantified different ATIs in few cultivars across different wheat species. These authors have shown a wide variability in ATI content and composition across the investigated cultivars. Consequently, there is the potential

Fig. 4 Fine-mapping of the major QTL (2275974D) for ATI 0.19-like on chromosome 3B. Genes detected in the target region 75–80 Mbp on chromosome 3B are marked by black boxes, while the two potential candidate genes TraesCS3B02G111100 and TraesCS3B02G111200 are marked by red and green boxes, respectively. Linkage disequilibrium (r^2) among all the markers in the region 75–80 Mbp is shown in the heatmap



to change contents and composition of ATIs in wheat by selecting specific cultivars.

The success of this choice along the wheat supply chain, however, depends besides other factors on the extent to how much the expression of the single ATI proteins is influenced by the environmental conditions, where the cultivars are grown. In particular, only traits with high heritability (at least 0.6, and larger) were stably expressed across different environments and growing conditions and can therefore be manipulated across supply chains via choice of cultivars. For six out of eight ATIs, we determined medium to low heritability values (Table 1), showing the strong environmental impact. By contrast, for ATI 0.28 and ATI CM2, heritabilities of 0.70 and 0.79 were obtained, respectively. For commonly investigated quality traits that are important for milling and bread making quality, heritability values

ranged from 0.65 for protein content up to almost 0.9 for sedimentation volume, hectoliter weight and thousand kernel weight measured on the same samples (Rapp et al. 2018). This underlines clearly that most of ATI proteins seem to be strongly influenced by environmental conditions as was also demonstrated by Prandi et al. (2013) and are therefore difficult to be manipulated along global wheat supply chains via choice of cultivar. Interestingly, the high heritability values obtained for ATI 0.28 and ATI CM2 confirm results from Geisslitz et al. (2020). By contrast, our heritability estimates for the other ATIs, especially 0.19, CM1 and CM3, are lower than the values reported by Geisslitz et al. (2020). As estimates of variance components and heritabilities have large errors, and as Geisslitz et al. (2020) used only eight varieties for bread wheat, we speculate that their heritability values might have been overestimated. Thus, the amount of

Table 4 List of eight candidate genes identified for ATI 0.19-like, ATI CM1, ATI CM2 and total ATI content through functional analysis on all mapped QTL explaining more than 10% of genotypic variance

Trait	GeneID	Pfam ^a	InterPro ^a	Chr	start	end
<i>ATI 0.19-like</i>						
	TraesCS3B02G111100	PF00234: Protease inhibitor/seed storage/LTP family	IPR006105: Cereal seed allergen/trypsin and alpha-amylase inhibitor, conserved site; IPR006106: Cereal seed allergen/grain softness/trypsin and alpha-amylase inhibitor; IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain; IPR016309: Alpha-amylase inhibitor/seed allergen	chr3B	77,315,338	77,316,027
	TraesCS3B02G111200	PF00234: Protease inhibitor/seed storage/LTP family	IPR006105: Cereal seed allergen/trypsin and alpha-amylase inhibitor, conserved site; IPR006106: Cereal seed allergen/grain softness/trypsin and alpha-amylase inhibitor; IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain; IPR016309: Alpha-amylase inhibitor/seed allergen	chr3B	77,349,935	77,350,623
	TraesCS3B02G294800	PF14368: Probable lipid transfer	IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain	chr3B	4.74E+08	4.74E+08
<i>ATI CM1</i>						
	TraesCS1A02G048700	PF14368: Probable lipid transfer	IPR000528: Plant lipid transfer protein/Par allergen; IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain	chr1A	30,196,596	30,197,675
<i>ATI CM2</i>						
	TraesCS7B02G072000	PF00234: Protease inhibitor/seed storage/LTP family	IPR006105: Cereal seed allergen/trypsin and alpha-amylase inhibitor, conserved site; IPR006106: Cereal seed allergen/grain softness/trypsin and alpha-amylase inhibitor; IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain;	chr7B	79,308,131	79,309,947
	TraesCS1D02G163900	PF14368: Probable lipid transfer	IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain;	chr1D	2.35E+08	2.35E+08
	TraesCS1D02G164000	PF14368: Probable lipid transfer	IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain;	chr1D	2.35E+08	2.35E+08
Total ATI	TraesCS3B02G170800	PF00234: Protease inhibitor/seed storage/LTP family	IPR000528: Plant lipid transfer protein/Par allergen; IPR016140: Bifunctional inhibitor/plant lipid transfer protein/seed storage helical domain	Chr3B	1.72E+08	1.72E+08

^a Protein families database

ATI 0.28 and ATI CM2 might be influenced quite successfully along supply chains by choosing cultivars with low ATI 0.28 and CM2 content. However, total ATI content will most likely not change much, since these two ATIs contributed only 4–13% to the total ATI content. Furthermore, Call et al. (2020) have only shown an intermediate correlation coefficient of $r=0.61$ between ATI content and their trypsin inhibitory activity. Hence, more research is needed on biological relevant activities and allergenicity of different ATI proteins as well as possibilities to manipulate these activities along the wheat supply chain through agronomic and other technological processes. Very recent studies investigated the contribution of food processing in ATI degradation by using different sourdough fermentations with lactic acid bacteria (Fraberger et al. 2020), or by comparing different bread making processes (Huang et al. 2020). Both studies showed the ability of sourdough fermentation to degrade ATI proteins for more than 22%. Moreover, Huang et al. showed a significant decrease in pro-inflammatory bioactivity of ATI underlining the potential of the bread making process on ATI.

Breeding wheat cultivars for lower ATI content

Selecting between registered wheat cultivars those with appropriate ATI profiles might be seen as a potential strategy to reduce allergenic and pro-inflammatory ATI content. Thus, targeted wheat breeding might contribute to reduce NCWS. Prerequisites for a successful breeding for low ATI content are an existing genetic variance within the elite wheat germplasm, high heritabilities and no negative correlations with important agronomic and quality traits. As already discussed above, we determined a large genetic variance for all the measured eight ATI proteins but higher heritabilities only for ATI 0.28 and ATI CM2. Notably, correlation coefficients with important quality traits in wheat were for all ATIs close to zero (Table 2), indicating that breeding for lower ATIs would not hinder selection for baking quality. Thus, wheat breeding could contribute to a reduction in ATI content, but selection gain is limited for most ATI proteins except ATI 0.28 and ATI CM2, due to low heritability for the other ATIs. Furthermore, the recent methodology to determine ATI content is work and cost intensive as well as too slow for a high sample throughput, which is needed in intermediate stages of large wheat breeding programs. Thus, the development of a fast method, such as a simple enzyme linked immune assay (ELISA), would be crucial warranting further research.

Interestingly, we determined also low correlation coefficients between ATI contents and protein content determined by NIRS (ICC standard method 159, ICC, Vienna, Austria), which confirms another recent study (Geisslitz et al. 2020). Consequently, a high protein content does not automatically imply a high ATI content that for now needs to be measured

independently for all existing ATI proteins. Surprisingly, correlation coefficients between ATI proteins and falling number were close to zero (Table 2). The falling number is widely applied in wheat supply chains as an indirect method to determine the α -amylase activity. Similarly, Call et al. (2020) have shown only an intermediate correlation coefficient between ATI content and trypsin inhibitory activity. Consequently, more research is needed on the role and effects of ATI in cereals and cereal products.

The 149 investigated wheat cultivars of our study were registered from 1921 to 2013 and originated from different European countries. We therefore investigated the possible existence of temporal or regional trends in the ATI content of different wheat cultivars. For all eight ATI proteins as well as total ATI content, there was neither a temporal (Fig. 2) nor a regional trend (Fig. S1), showing that wheat breeders did not select for changes in ATI contents in Central and Eastern Europe in the last century, neither directly nor indirectly via correlation with other traits. This confirms our experience from many discussions with breeder colleagues and is also in accord with a study that found no significant difference between old and modern Austrian wheat evaluated for total ATI content (Shewry et al. 2020). This is in contrast to statements that modern bread wheat cultivars have increased ATI content compared to old varieties as originally hypothesized by Junker et al. (2012), and now widely stated in popular press and social media.

Furthermore, the absence of a trend over time and geographic regions regarding ATI content is also remarkable, since it is often claimed that breeders have actively enhanced ATIs in wheat due to insect resistance breeding. Beside the lacking trend of increased ATI content in modern wheat cultivars, wheat breeders have yet not invested much in insect resistance breeding except for orange blossom midge resistance (McKenzie et al. 2002), which has received increased attention within the last years. Amylase (trypsin) inhibitors naturally exist in wheat and might inhibit proteinases from insects such as weevils. Several early studies (Buonocore et al. 1980; Ryan 1990; Carbonero et al. 1999) reported amylase inhibitors to be active against storage pests such as *Tenebrio molitor*. However, to the best of our knowledge, the functions of ATIs in wheat are not yet sufficiently explored and warrant further research. Summarizing, to date, wheat breeding has not altered ATI contents and composition of wheat cultivars.

ATI proteins are influenced by many small and few major QTL

The genome-wide association mapping identified between four (ATI CM1, ATI CM17) to 13 QTL (ATI CM16) for the eight investigated ATIs (Table 3, Table S2). Many of these QTL explained only a small proportion of the genotypic

variance, which is common for quantitatively inherited traits. However, for each ATI at least one major QTL explaining > 23% of the genotypic variance was identified with two very large-effect QTL on chromosomes 6B and 3B explaining 68.7% and 70.6% of the genotypic variance for ATI 0.28 and ATI 0.19-like, respectively. We therefore tried to identify candidate genes with functional annotations similar to the different domains of ATIs based on the reference sequence (IWGSC RefSeq v1.0). We searched for all 30 QTL which explained > 10% of the genotypic variance, but could only determine eight potential candidate genes for QTL of ATI 0.19-like, ATI CM1, ATI CM2, and for total ATI content (Table 4). Notably, all these loci encoded wheat proteins of the lipid transfer protein family that are structurally related to ATI proteins (Asero et al. 2001). For the very large-effect QTL explaining 68.7% of the genotypic variance of ATI 0.28, we could not identify any gene in that chromosomal region with functional annotations similar to an ATI (Fig. S3). The chromosomal region of the major QTL detected for ATI 0.19-like harbors two potential candidate genes TraesCS3B02G111100 and TraesCS3B02G111200 (Fig. 4). However, a LD breakage within the peak region prevents clear conclusions. Thus, further research is needed to clarify the genetic architecture of ATI proteins in more detail across wider germplasm or within biparental mapping families.

The same chromosomal groups 3 and 6 were reported to mainly influence dimeric (0.19) and monomeric (0.28) ATIs, respectively. Earlier studies have reported that dimeric ATIs including ATI 0.19 were mapped to 3B (Sanchez-Monge et al. 1986; Singh and Skerritt 2001; Juhász et al. 2018). Similarly, ATI 0.28 was reported to be encoded by a single chromosomal region on 6D (Sanchez-Monge et al. 1986; Singh and Skerritt 2001; Bose et al. 2020). However, the mapping approach was different compared to our study and no information related to exact chromosomal location has been reported. Therefore, it is quite difficult to align the results from those studies to the current one. In contrast to the prior results, we additionally found further QTL for both ATIs on different chromosomes (Table S2). Furthermore, for the remaining six ATI proteins and total ATI content, we identified only medium to small effect QTL in different chromosomal locations in addition to the previously reported chromosomes group 4 and 7 for CM type ATIs (Fra-Mon et al. 1984; Salcedo et al. 1984). Consequently, the genetic architecture of ATI proteins appears more complex than reported and requires further research.

We identified high positive correlation coefficients for the expression of CM type ATIs (Table 2). Consequently, several CM type ATIs were controlled by co-localized QTL. A recent proteome study confirmed the high similarity of some ATIs including CM16 and CM17 by detecting common peptides after trypsin digestion (Geisslitz et al. 2020).

Cluster analysis of ATI-like proteins by mean of discovery proteomics on wheat extracted proteins revealed that CM proteins fall into the same cluster (Bose et al. 2020). This makes physiological sense, since CM ATIs form stable non-covalently tetramers with two molecules of CM3 associating with two other CM ATIs (Gomez et al. 1989; Altenbach et al. 2011). Thus, the association within CM type ATI proteins seems systematic and further research is needed to elucidate this association in more detail. For instance, whether this association results from linked or pleiotropic genes, or similar pathways of gene or post-translational regulation.

Overall, the findings on the genetic architecture of ATI proteins can facilitate targeted wheat breeding. First, single marker assays could be developed for the major QTL on chromosome 3B and 6B for ATI 0.19-like and ATI 0.28, respectively. As both alleles at these QTL are widely present in modern cultivars (data not shown), this would allow an easy fixation of the favorable alleles across early generation selection. Secondly, we determined medium to high prediction abilities for all ATI proteins and total ATI content from genomic selection using either marker-assisted selection based only on QTL explaining > 10% of the genotypic variance or weighted RR-BLUP (Fig. S4). These prediction abilities are comparable with those of other traits in wheat, such as asparagine content, protein content and grain yield (Rapp et al. 2018; Michel et al. 2019) and high enough to increase annual selection gain by use of genomic selection (Marulanda et al. 2016). An efficient application in wheat breeding would require further research extending genomic calibrations on a wider range of cultivars coming from different breeding programs. However, starting to implement low ATI content as new trait in wheat breeding is an expensive and long-term decision with limited success due to the low heritability of many ATI proteins. Thus, we would only recommend it if medical and market necessities are further confirmed and if other alternatives such as milling and bread making technologies turn out to be less efficient than breeding in reducing ATI content along wheat supply chains. On the other hand, gene editing technologies might be another alternative to decrease content and unwanted biological activities of ATIs (Camerlengo et al. 2020; Kalunke et al. 2020). However, this is applicable only if target ATIs are encoded by major loci which appear not to be the case for most of them according to the current study.

Conclusions

Based on the evaluation of eight major ATI proteins in a diverse collection of wheat cultivars, we could clearly show that wheat breeding has not altered the content and composition of ATI proteins neither directly nor indirectly in the last century of breeding history in Central Europe. We identified

large differences in the content and composition of the ATI proteins in the different cultivars, and heritability values were high only for ATI 0.28 and ATI CM2. Thus, these two proteins bear the potential to be successfully manipulated via varietal choice across international wheat supply chains, but they contributed only up to 13% of total ATI content in our study. The genetic architecture of ATI proteins appears complex with many minor but few major QTL with large effect. Based on the wheat reference sequence, we could only identify eight potential candidate genes behind all major QTL that warrant future research. Additionally, further research is needed to validate the impact of ATI proteins that are major allergens, also on other aspects of human health. Furthermore, rapid methods must be elaborated to quantify either ATI proteins or determine ATI pro-inflammatory bioactivity in cereals or food samples within a short time. Eventually, the possibility to reduce ATI content and activity in wheat products must then be extended along the wheat supply chain to milling and bread or pasta making technology, including different dough fermentations or the use of enzymes to inactivate ATIs in wheat products in order to elaborate more efficient methods than selecting cultivars for low heritable traits.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-021-03906-y>.

Acknowledgements The project was supported by the German Research Foundation projects DFG LO 1816/4-1, TE599/3-1 SCHU 646/17-1 (to FL, ST, DS) and a research grant from the Leibniz Foundation (WheatScan, SAW-2016-DFA-2) to DS. KELH was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program (BETTERWHEAT–FKZ: 2818405A18).

Author contribution statement CFHL, ST and DS designed the study; MS collected data; VC and MN performed and standardized the wheats extractions; KELH, TW and WL performed the analyses; KELH, MS, TW, WL, DS and TS wrote the paper; all coauthors approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The authors declare that the experiments do not involve humans or animals and comply with the current laws in Germany.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alaux M, Rogers J, Letellier T, Flores R, Alfama F, Pommier C, Mohellibi N, Durand S, Kimmel E, Michotey C, Guerche C, Loaec M, Lainé M, Steinbach D, Choulet F, Rimbert H, Leroy P, Guilhot N, Salse J, Feuillet C, International Wheat Genome Sequencing Consortium, Paux E, Eversole K, Adam-Blondon AF, Quesneville H (2018) Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biology* 19:111
- Altenbach SB, Vensel WH, Dupont FM (2011) The spectrum of low molecular weight alpha-amylase/protease inhibitor genes expressed in the US bread wheat cultivar Butte 86. *BMC Res Notes* 4:242
- Asero R, Mistrello G, Roncarolo D, de Vries SC, Gautier MF, Ciurana CL, Verbeek E, Mohammadi T, Knul-Brettlova V, Akkerdaas JH, Bulder I, Aalberse RC, van Ree R (2001) Lipid transfer protein: a pan-allergen in plant-derived foods that is highly resistant to pepsin digestion. *Int Arch Allergy Immunol*. <https://doi.org/10.1159/000053671>
- Ashfaq-Khan M, Aslam M, Qureshi MA, Senkowski MS, Yen-Weng S, Strand S, Kim YO, Pickert G, Schantenberg J, Schuppan D (2019) Dietary wheat amylase trypsin inhibitors promote features of murine non-alcoholic fatty liver disease. *Sci Rep* 9:17463
- Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23(10):1294–1296
- Bellinghausen I, Weigmann B, Zevallos V, Maxeiner J, Reissig S, Waisman A, Schuppan D, Saloga JMD (2019) Wheat amylase-trypsin inhibitors exacerbate intestinal and airway allergic immune responses in humanized mice. *J Allergy Clin Immunol* 143(1):201–212
- Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods* 2(8):587–589
- Bose U, Byrne K, Howit C, Colgrave ML (2019) Targeted proteomics to monitor the extraction efficiency and levels of barley α -amylase trypsin inhibitors that are implicated in non-coeliac gluten sensitivity. *J Chromatogr A* 1600:55–64
- Bose U, Juhasz A, Broadbent JA, Byrne K, Colgrave HCA, ML, (2020) Identification and quantitation of amylase trypsin inhibitors across cultivars representing the diversity of bread wheat. *J Proteome Res*. <https://doi.org/10.1021/acs.jproteome.0c00059>
- Buonocore V, Gramenzi F, Pace W, Petrucci T, Poerio E, Silano V (1980) Interaction of wheat monomeric and dimeric protein inhibitors with α -Amylase from yellow mealworm (*Tenebrio molitor* L.Larva). *J Biochem* 187:637–645
- Call L, Kapeller M, Grausgruber H, Reiter E, Schoenlechner R, D'Amico S (2020) Effects of species and breeding on wheat protein composition. *J Cereal Sci*. <https://doi.org/10.1016/j.jcs.2020.102974>
- Camerlengo F, Frittelli A, Sparks C, Doherty A, Martignago D, Larré C, Lupi R, Sestili F, Masci S (2020) CRISPR-Cas9 multiplex

- editing of the α -amylase/trypsin inhibitor genes to reduce allergen proteins in durum wheat. *Front Sustain Food Syst* 4:104
- Caminero A, Verdu EF (2019) Metabolism of wheat proteins by intestinal microbes: implications for wheat related disorders. *Gastroentol y Hepatol* 42(7):449–457
- Carbonero P, Diaz I, Vicente-Carbajosa J, Alfonso-Rubi J, Gaddour K, Lara P (1999) Cereal α -amylase/trypsin inhibitors and transgenic insect resistance. In: Mugnozza GTS, Porceddu E, Pagnotta MA (eds) *Genetics and Breeding for Crop Quality and Resistance*, 1st edn. Springer, Dordrecht, pp 147–158
- Carroccio A, Mansueto P, D'Alcamo A, Iacono G (2013) Non-celiac wheat sensitivity as an allergic condition: personal experience and narrative review. *Am J Gastroenterol* 108(12):1845–1852
- Catassi C, Alaedini A, Bojarski C, Bonaz B, Bouma G, Carroccio A, Castillejo G, De Magistris L, Dieterich W, Di Liberto D, Elli L, Fasano A, Hadjivassiliou M, Kurien M, Lionetti E, Mulder CJ, Rostami K, Sapone A, Scherf K, Schuppan D, Trott N, Volta U, Zevallos V, Zopf Y, Sanders DS (2017) The overlapping area of non-celiac gluten sensitivity (NCGS) and wheat-sensitive irritable bowel syndrome (IBS): an update. *Nutrients* 9(11):1268
- Cuccioloni M, Mozzicafreddo M, Bonfili L, Cecarini V, Giangrossi M, Falconi M, Saitoh SI, Eleuteri AM, Angeletti M (2017) Interfering with the high-affinity interaction between wheat amylase trypsin inhibitor CM3 and toll-like receptor 4: in silico and biosensor-based studies. *Sci Rep* 7:13169
- Dinu M, Whittaker A, Pagliai G, Benedettelli S, Sofi FMD (2018) Ancient wheat species and human health: biochemical and clinical implications. *J Nutr Biochem* 52:1–9
- Distler U, Lacki MK, Schumann S, Wanninger M, Tenzer S (2019) Enhancing sensitivity of microflow-based bottom-up proteomics through postcolumn solvent addition. *Anal Chem* 91(12):7510–7515
- Dos Santos GM, Zevallos VF, Pesí A, Stoye NM, Nguyen VTT, Radyushkin K, Schwiertz A, Schmitt U, Schuppan D, Endres K (2020) Dietary wheat amylase trypsin inhibitors impact Alzheimer's disease pathology in 5xFAD model mice. *Int J Mol Sci* 21(17):6288
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome J* 4:250–255
- Fasano A, Sapone A, Zevallos V, Schuppan D (2015) Nonceliac gluten sensitivity. *Gastroenterology* 148(6):1195–1204
- Food and agriculture organization of the united nations (2019) *Food outlook - Biannual Report on Global Food Markets – Novemb 2019*. Rome.
- Fraberger V, Ladurner M, Nemeč A, Grunwald-Gruber C, Call LM, Hochegger R, Domig KJ, D'Amico S (2020) Insights into the potential of sourdough-related lactic acid bacteria to degrade proteins in wheat. *Microorganisms* 8(11):1689
- Fra-Mon P, Salcedo G, Aragoncillo C, Garcia-Olmedo F (1984) Chromosomal assignment of genes controlling salt-soluble proteins (albumins and globulins) in wheat and related species. *Theor Appl Genet* 69:167–172
- Franco O, Rigden DJ, Melo FR, Grossi-de-Sa MF (2002) Plant α -amylase inhibitors and their interaction with insect α -amylases. Structure, function and potential for crop protection. *Eur J Biochem* 269:397–412
- Fritscher-Ravens A, Schuppan D, Ellrichmann M, Schoch S, Röcken C, Brasch J, Bethge J, Böttner M, Klose J, Milla PJ (2014) Confocal endomicroscopy shows food-associated changes in the intestinal mucosa of patients with irritable bowel syndrome. *Gastroenterology* 147(5):1012–20.e4
- Fritscher-Ravens A, Pflaum T, Mösinger M, Ruchay Z, Röcken C, Milla PJ, Das M, Böttner M, Wedel T, Schuppan D (2019) Many patients with irritable bowel syndrome have atypical food allergies not associated with immunoglobulin E. *Gastroenterology* 157(1):109–118.e5
- Geisslitz S, Ludwig C, Scherf KA, Koehler P (2018) Targeted LC-MS/MS reveals similar contents of α -amylase/trypsin-inhibitors as putative triggers of nonceliac gluten sensitivity in all wheat species except einkorn. *J Agri Food Chem* 66:12395–12403
- Geisslitz S, Longin CFH, Koehler P, Scherf KA (2020) Comparative quantitative LC-MS/MS analysis of 13 amylase/trypsin inhibitors in ancient and modern *Triticum* species. *Sci Rep* 10:14570
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) *ASReml user guide release 3.0*. VSN International Ltd, Hemel Hempstead, UK
- Gomez L, Sanchez-Monge R, Garcia-Olmeda F, Salcedo G (1989) Wheat tetrameric inhibitors of insect α -amylases: Allopolyploid heterosis at the molecular level. *Proc Nat Acad Sci* 86:3242–3246
- Huang X, Schuppan D, Rojas Tovar LE, Zevallos VF, Loponen J, Gänzle M (2020) Sourdough fermentation degrades wheat α -amylase/trypsin inhibitor (ATI) and reduces pro-inflammatory activity. *Foods* 9:943
- Juhász A, Belova T, Florides CG et al (2018) Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Sci Adv*. <https://doi.org/10.1126/sciadv.aar8602>
- Junker YY, Zeissig S, Kim S-J, Barisani D, Wieser H, Leffler DA, Zevallos V, Libermann TA, Dillon S, Freitag TL, Kelly CP, Schuppan D (2012) Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. *J Exp Med* 209:2395–2408
- Kalunke RM, Tundo S, Sestili F, Camerlengo F, Lafiandra D, Lupi R, Larré C, Denery-Papini S, Islam S, Ma W, D'Amico S, Masci S (2020) Reduction of allergenic potential in bread wheat RNAi transgenic lines silenced for CM3, CM16 and 0.28 ATI genes. *Int J Mol Sci* 21(16):5817
- Li H, Vikram P, Singh RP et al (2015) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genom* 16:216
- Marulanda JJ, Mi X, Melchinger AE, Xu J-L, Würschum T, Longin CFH (2016) Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. *Theor Appl Genet* 129:1901–1913
- McKenzie RIH, Lamb RJ, Aung T, Wise IL, Barker P, Olfert OO (2002) Inheritance of resistance to wheat midge, *Sitodiplosis mosellana*, in spring wheat. *Plant Breed* 121:383–388
- Michel S, Loschenberger F, Ametz C, Pachler B, Sperry E, Burstmayr H (2019) Simultaneous selection for grain yield and protein content in genomics-assisted wheat breeding. *Theor Appl Genet* 132:1745–1760
- Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, Myles S (2015) LinkImpute: fast and accurate genotype imputation for non-model organisms. *G3 Genes Genomes Genet* 5:2383–2390
- Oda Y, Matsunaga T, Fukuyama K, Miyazaki T, Morimoto T (1997) Tertiary and quaternary structures of 019 α -amylase inhibitor from wheat kernel determined by X-ray analysis at 206 Å resolution. *Biochemistry* 36:13503–13511
- Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaíno JA (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47(D1):D442–D450
- Pickert G, Wirtz S, Matzner J et al (2020) Wheat consumption aggravates colitis in mice via amylase trypsin inhibitor-mediated dysbiosis. *Gastroenterology* 159(1):257–272
- Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177:1881–1888

- Prandi B, Faccini A, Tedeschi T, Galaverna G, Sforza S (2013) LC/MS analysis of proteolytic peptides in wheat extracts for determining the content of the allergen amylase/trypsin inhibitor CM3: Influence of growing area and variety. *Food Chem* 140:141–146
- Pronin D, Borner A, Scherf KA (2020) Old and modern wheat (*Triticum aestivum* L.) cultivars and their potential to elicit celiac disease. *Food Chem* 339:127952
- Pronin D, Borner A, Weber H, Scherf KA (2020b) Wheat (*Triticum aestivum* L.) breeding from 1891 to 2010 contributed to increasing yield and glutenin contents but decreasing protein and gliadin contents. *J Agri Food Chem* 68:13247–13256
- R Development Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Rapp M, Schwadorf K, Leiser WL, Wursschum T, Longin CFH (2018) Assessing the variation and genetic architecture of asparagine content in wheat: what can plant breeding contribute to a reduction in the acrylamide precursor? *Theor Appl Genet* 131:2427–2437
- Reig-Otero Y, Mañes J, Manyes L (2018) Amylase–trypsin inhibitors in wheat and other cereals as potential activators of the effects of nonceliac gluten sensitivity. *J Med Food* 21:207–214
- Ryan CA (1990) Protease inhibitors in plants: genes for improving defenses against insects and pathogens. *Annu Rev Phytopatol* 28:425–429
- Salcedo G, Fra-Mon P, Molina-Cano JL, Aragoncillo C, Garcia-Olmedo F (1984) Genetics of CM-proteins (A-hordeins) in barley. *Theor Appl Genet* 68:53–59
- Salcedo G, Quirce S, Diaz-Perales A (2011) Wheat allergens associated with Baker's asthma. *J Investig Allergol Clin Immunol* 21(2):81–92
- Sanchez-Monge R, Barber D, Mendez E, Garcia-Olmedo F, Salcedo G (1986) Genes encoding α -amylase inhibitors are located in the short arms of chromosomes 3B, 3D and 6D of wheat (*Triticum aestivum* L.). *Theor Appl Genet* 72:108–113
- Schuppan D, Gisbert-Schuppan K (2019) Wheat Syndromes: How Wheat Gluten and ATI Cause Disease. Springer Nature, London
- Schuppan D, Zevallos V (2015) Wheat amylase trypsin inhibitors as nutritional activators of innate immunity *Digestive. Diseases* 33:260–263
- Schuppan D, Pickert G, Ashfaq-Khan M, Zevallos V (2015) Non-celiac wheat sensitivity: differential diagnosis, triggers and implications. *Best Pract Res Clin Gastroenterol* 29(3):469–476
- Shewry PR (2009) Wheat. *J Exp Bot* 60:1537–1553
- Shewry PR, Hassall KL, Grausgruber H, Andersson AAM, Lampi AM, Piironen V, Rakszegi M, Ward JL, Lovegrove A (2020) Do modern types of wheat have lower quality for human health? *Nutr Bull* 45(2):362–373
- Sielaff M, Curella V, Neerukonda M, Afzal M, Hassouni EL, K, Disler Ute, Schuppan D, Longin CFH, Tenzer S, (2021) 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(3):1544–1557
- Silano V, Pochiari F, Kasarda DD (1973) Physical characterization of α -amylase inhibitors from wheat. *Biochem Biophys Acta* 317:139–148
- Silano V, Poerio E, Buonocore V (1977) A model for the interaction of wheat monomeric and dimeric protein inhibitors with α -amylase. *Mol Cell Biochem* 18(2–3):87–91
- Singh J, Skerritt JH (2001) Chromosomal control of albumins and globulins in wheat grain assessed using different fractionation procedures. *J of Cereal Science* 33:163–181
- The International Wheat Genome Sequencing Consortium (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361:eaar7191
- Tundo S, Lupi R, Lafond M, Giardina T, Larré C, Denery-Papini S, Morisset M, Kalunke R, Sestili F, Masci S (2018) Wheat ATI CM3, CM16 and 028 allergens produced in *pichia pastoris* display a different eliciting potential in food allergy to wheat. *Plants* 7(4):101
- Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* 154:1839–1849
- Wei T, Simko V (2017) R package “corrplot”: Visualization of a Correlation Matrix (Version 084). R Foundation for Statistical Computing, Vienna, Austria
- Williams E, Piepho HP, Whitaker D (2011) Augmented p-rep designs. *Biometrical J* 53:19–27
- Würschum T, Leiser WL, Kazman E, Longin CFH (2016) Genetic control of protein content and sedimentation volume in European winter wheat cultivars. *Theor Appl Genet* 129:1685–1696
- Zevallos VF, Raker V, Tenzer S, Jimenez-Calvente C, Ashfaq-Khan M, Russel N, Pickert G, Schild H, Steinbrink K, Schuppan D (2017) Nutritional wheat amylase-trypsin inhibitors promote intestinal inflammation via activation of myeloid cells. *Gastroenterology* 152(5):1100–1113.e1112
- Zevallos VF, Raker VK, Maxeiner J, Schiltes P, Steinbrink K, Schuppan D (2019) Dietary wheat amylase trypsin inhibitors exacerbate murine allergic airway inflammation. *Eur J of Nutrition* 58:1507–1514
- Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity* 112:638–645

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