Increased Wheat Protein Content *via* Introgression of an HMW Glutenin Selectively Reshapes the Grain Proteome

Authors

Hui Cao, Owen Duncan, Shahidul Islam, Jingjuan Zhang, Wujun Ma, and A. Harvey Millar

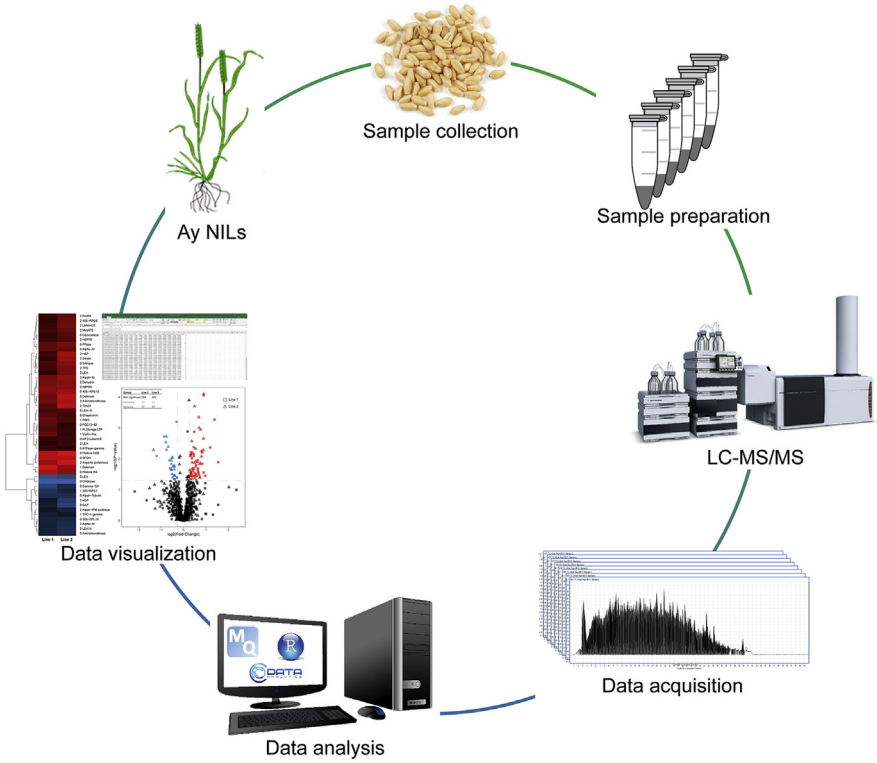
Correspondence

W.Ma@murdoch.edu.au; harvey.millar@uwa.edu.au

In Brief

The differences between the grain proteome of three wheat cultivars and corresponding Ay HMW-GS–introgressed near-isogenic lines have been determined. In addition to increased abundance of 1Ay HMW-GS, 115 differentially expressed proteins were also discovered. This revealed that introgression of the 1Ay21* HMW-GS increases wheat grain protein content and improves bread-making quality in association with a wider reshaping of the grain proteome network.

Graphical Abstract



Highlights

- Ay HMW-GS itself only contributes to 20% of the significant GPC increase in Ay NILs.
- Ay HMW-GS enhances other storage protein and protein synthesis machinery abundances.
- Expression of genes encoding Ay HMW-GS–induced proteins are strongly coexpressed.
- It provides a mechanistic model to influence future wheat quality breeding programs.

2021, Mol Cell Proteomics *20*, 100097

https://doi.org/10.1016/j.mcpro.2021.100097

Increased Wheat Protein Content *via* Introgression of an HMW Glutenin Selectively Reshapes the Grain Proteome

Hui Cao[1,2], Owen Duncan[1,2], Shahidul Islam[3,4], Jingjuan Zhang[3,4], Wujun Ma[3,4,*], and A. Harvey Millar[1,2,*]

Introgression of a high-molecular-weight glutenin subunit (HMW-GS) allele, *1Ay21, into commercial wheat cultivars increased overall grain protein content and bread-making quality, but the role of proteins beyond this HMW-GS itself was unknown. In addition to increased abundance of 1Ay HMW-GS, 115 differentially accumulated proteins (DAPs) were discovered between three cultivars and corresponding introgressed near-isogenic lines. Functional category analysis showed that the DAPs were predominantly other storage proteins and proteins involved in protein synthesis, protein folding, protein degradation, stress response, and grain development. Nearly half the genes encoding the DAPs showed strong coexpression patterns during grain development. Promoters of these genes are enriched in elements associated with transcription initiation and light response, indicating a potential connection between these *cis*-elements and grain protein accumulation. A model of how this HMW-GS enhances the abundance of machinery for protein synthesis and maturation during grain filling is proposed. This analysis not only provides insights into how introgression of the *1Ay21** improves grain protein content but also directs selection of protein candidates for future wheat quality breeding programs.**

Hexaploid wheat (*Triticum aestivum* L.; 2n = 6x = 42, AABBDD) is cultivated worldwide because of its value as a staple food and protein source. Apart from the relatively high starch content (60–70% of the whole-grain dry weight), the wheat grain protein content (GPC) varies from 8 to 15%, and this determines bread-making qualities (1). The GPC largely consists of storage proteins including glutenins and gliadins, of which glutenins determine the dough viscoelasticity and elasticity while gliadins contribute to dough extensibility (2, 3). Glutenins can be further divided into two groups, namely high-molecular-weight glutenin subunits (HMW-GSs, 70–140 kDa) and low-molecular-weight glutenin subunits (30–80 kDa) (4, 5).

Generally, HMW-GS encoding genes are located at the *Glu-1* loci on the long arms of chromosomes 1A, 1B, and 1D. Each *Glu-1* locus consists of two tightly linked genes with one gene encoding an x-type small subunit and the other gene encoding a y-type large subunit, named according to molecular weight differences between the encoded HMW-GS proteins (6, 7). Among these six subunits, three subunits (Glu-1Bx, Glu-1Dx, and Glu-1Dy) are always expressed in bread wheat cultivars, while two subunits (Glu-1Ax and Glu-1By) are sometimes expressed, but the Glu-1Ay subunit is expressed rarely (8). As a result, most common bread wheat cultivars usually have 3 to 5 expressed HMW-GSs. However, the *Glu-1Ay* gene has been found to be functionally expressed in wild diploid and tetraploid wheats (9, 10).

The GPC and composition are important traits in global bread wheat breeding programs because of the continuous growth in demand for specified industrial end use (11, 12). Glutenin, in particular, has become one of the main breeding targets because of its critical role in determining bread-making quality. For example, it has been previously reported that allelic variation of HMW-GS contributes to about 45 to 70% of the variation in bread-making quality, although it only contributes to about 10 to 12% of the total GPC (13–15). Furthermore, y-type HMW-GSs are generally considered to be more valuable for dough quality improvement than x-type HMW-GSs because they are longer and possess more cysteine residues for forming intermolecular and intramolecular disulfide bonds during dough development (3). Genetic engineering strategies have been widely used to improve the protein content and composition in wheat breeding, which includes the use of conventional cross-breeding processes to integrate additional genes into

From the [1]ARC Centre of Excellence in Plant Energy Biology, [2]School of Molecular Science, University of Western Australia, Crawley, Western Australia, Australia; [3]State Agricultural Biotechnology Centre, College of Science Health Engineering and Education, Murdoch University, Murdoch, Western Australia, Australia; [4]Australia-China Joint Centre for Wheat Improvement, Murdoch University, Perth, Western Australia, Australia

*For correspondence: A. Harvey Millar, harvey.millar@uwa.edu.au; Wujun Ma, W.Ma@murdoch.edu.au.

https://doi.org/10.1016/j.mcpro.2021.100097

background genotypes to produce extra protein functionality and increase the protein content (12). Margiotta *et al.* (1996) reported the identification of active *1Ax* and *1Ay* genes in several Swedish bread wheat lines, and a year later, Rogers *et al.* (1997) demonstrated that 1Ay HMW-GS had positive effects on bread-making quality, such as improved dough stability during mixing and enhanced gluten strength (16, 17). Since then, attention has been drawn to understanding the biological function of 1Ay HMW-GS and to estimate its value in wheat breeding programs. Li *et al.* (2006) (18) reported the investigation of HMW-GS variation among 205 cultivated emmer accessions, which led to the successful identification of novel 1Ay subunits and the recommendation of emmer accessions as a valuable genetic resource for quality improvement of common wheat. Another study, focusing on genetic variation of HMW-GSs in 1051 accessions from 13 *Triticum* subspecies including diploid, tetraploid, and hexaploid wheat, identified the expression of 1Ay subunits in *Triticum urartu* and suggested these unique HMW-GS alleles could be further utilized through direct hybrid production for quality improvement of common wheat (19).

Recently, Roy *et al.* integrated an expressed 1Ay allele from an Italian line into Australian commercial cultivars using a conventional breeding approach and observed better dough rheological properties and bread-making quality without significant changes in agronomic traits. Improved quality traits were observed, including higher protein content, glutenin/gliadin ratio and unextractable polymeric protein proportion, stronger dough strength, and better water absorption (20, 21). It is well known that improving grain yield and GPC simultaneously is difficult because of their inverse correlation (22). However, the introgression of the 1Ay21* HMW-GS in the Lincoln cultivar background resulted in an increase in both grain yield and protein content in glasshouse and field conditions (21). Yu *et al.* (23) cloned and characterized the *Glu-1Ay* gene sequence identified in the Italian line and suggested that it has a close evolutionary relationship with Ay genes of *T. urartu*.

However, how the introgressed *Glu-1Ay* gene increases bread wheat protein content without changing other agronomic traits is still unknown. Here, we have undertaken a systematic proteome analysis for three Ay introgressed near-isogenic lines (NILs) in Australian cultivar backgrounds. Our findings provide new insights into which grain proteins are increased or decreased in abundance in introgressed lines and propose a common link between them.

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

The article contains three sets of analyses, namely, analysis of changes in abundance of components in the grain proteome network, gene coexpression profiles analysis, and gene *cis*-acting element enrichment analysis.

Wheat grains from the BC4F4 generation (backcrossed for four generations followed by self-pollination for three generations from the F1 seeds) of NILs containing the 1Ay21* HMW-GS gene were used in this study. All proteomics data presented were obtained from four biological replicates, and each biological replicate includes 20 randomly selected mature wheat grains. Mass tolerance of 20 ppm and 1% false discovery rate were applied for both label-free quantification (LFQ) and intensity-based absolute quantification (iBAQ) in MaxQuant searches. One-way ANOVA was used to define the differentially accumulated proteins (DAPs) between parental cultivars and corresponding Ay-integrated lines.

The gene expression of these DAPs in 177 independent samples from across 23 wheat varieties, 21 organs, and 37 growth stages was retrieved from Wheat Expression Browser (http://www.wheat-expression.com/). Averaged gene expression data of each sample were used in this study. Spearman's correlation analysis was performed to determine the gene coexpression level. Gene promoter region *cis*-acting elements were identified using PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The subsequent *cis*-acting enrichment comparison was conducted between common DAPs and non-DAPs as a background. Fisher's exact tests were performed to evaluate the significant difference. All data processing, statistical analysis, and visualization were performed using R programming language (version 3.5.1).

Plant Materials

The introgression of the expressed *Glu-1Ay* gene from an Italian wheat line N11 into Australian wheat cultivars EGA Gregory, EGA Bonnie Rock, and Yitpi was completed using conventional cross-breeding processes as previously reported (20, 21). The EGA Bonnie Rock NILs were previously published by Roy *et al.* (20). In short, NILs were developed in each cultivar till the BC4F4 generation. The NILs and their corresponding parental cultivars were grown at the experimental field station of the Department of Primary Industries and Regional Development at South Perth, Western Australia, in 2017 following a complete randomized design with three replicates. The experimental field station experiment was irrigated. The individual plot size was 0.6 m^2 and hand sowing was conducted to maintain uniform plant density. The parental cultivar and its two corresponding Ay-integrated lines for each of the three cultivars were selected in this study. Grains were harvested from the field plot at full maturity, and 100 g of subsamples of each replicate was prepared for phenotyping. The GPC and gluten abundance were measured using near-infrared transmission spectrophotometry through the "CropScan 3000F Flour and Grain Analyser" (Next Instruments). Each measurement was taken by averaging the data of ten individual scanning events.

Grain Protein Sample Preparation

Protein extraction involved very fine grinding of whole grains with excess reductant and a high SDS concentration to combine the extractable polymeric protein and unextractable polymeric protein fractions from wheat grains (24) into a single sample. Samples of four biological replicates for each line (supplemental Table S1*A*) and 20 seeds per replicate were ground with a mortar and a pestle under liquid nitrogen. 200 mg of finely ground extract was weighed for further total protein extraction using a chloroform/methanol precipitation approach (25). Briefly, samples were thoroughly mixed with 400-μl extraction buffer that contained 125 mM Tris HCl, pH 7.5, 7% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 0.5% (w/v) PVP40, and Roche protease inhibitor cocktail (one tablet per 50 ml of extraction buffer, Roche). After rocking on ice for 10 min, the mixture was centrifuged at 10,000*g* at 4 °C for 5 min to separate the supernatant and pellet. Two hundred microliter of the supernatant was then moved into a new tube followed by a precipitation step *via* adding 800-μl methanol, 200-μl

chloroform, and 500-µl distilled deionized water. After another centrifugation at 10,000*g* at 4 °C for 5 min, the upper aqueous phase was carefully removed and the protein pellet was washed twice with 500 µl methanol. The protein pellet was then incubated twice with 1 ml 90% (v/v) acetone at –20 °C for at least 1 h. After drying at room temperature, the dried protein pellet was resuspended with 40- to 100-µl resuspension buffer (50 mM ammonium bicarbonate, 1% SDS, and 10 mM DL-DTT). The amido black quantification method was used to determine the protein concentration (26).

To enhance protein digestion, 200 µg of protein was first incubated with 20 mM DL-DTT for 30 min in the dark, followed by a second incubation with 20 mM iodoacetamide for 30 min in the dark. After diluting the concentration of SDS to 0.1% (v/v), the protein was digested overnight using trypsin (Sigma) with a trypsin–protein ratio of 1:50. SDS removal was performed using the method reported by Yang *et al.* (27) with minor modifications. Briefly, peptide samples were injected into an off-line Agilent 1200 series HPLC configured with two J4SDS-2 guard columns (PolyLC, Columbia). Peptides were eluted from the columns by applying 60% (v/v) acetonitrile for 2 to 5 min and were collected from the eluate from 3.7 min to 4.5 min. After drying down in a vacuum centrifuge, dried peptides were resuspended with 5% (v/v) acetonitrile and 0.1% (v/v) formic acid in water. This was followed by a further filtering step using 0.22-µm centrifugal filter units (Millipore) to remove any insoluble materials from the samples. Purified peptide suspensions were stored at –80 °C for further use.

Acquisition of Mass Spectra

A fixed volume of 2 µl of purified peptide suspension from each sample was injected into a HPLC chip (Polaris-HR-Chip-3C18) using a capillary pump with a flow rate of 1.5 µl/min. Peptides were eluted from the column into an online Agilent 6550 Q-TOF MS (Agilent Technologies). A 2-h gradient generated by a 1200 series nano pump (Agilent Technologies) with a nano flow of 300 nl/min was used to separate peptides for MS. The elution gradient started with 5% (v/v) solution B (0.1% (v/v) formic acid in acetonitrile), increased to 6% in 6 min, and then 6 to 22% in 84 min, 22 to 35% in 5 min, 35 to 90% in 3 min, 90% in 4 min, and 90 to 5% in 2 min. Parameter settings for each mass spectrum acquisition were previously reported by Duncan *et al.*, (28). Briefly, a data-dependent mode and a scan range from 300 to 1750 mz were used for MS acquisition. MS data were collected at eight spectra per second, and MS/MS data were collected at four spectra per second. Ions were dynamically excluded for 6 s after fragmentation. In total, MS data for 36 samples were successfully collected.

LFQ of Precursor Ion Intensities

To obtain protein relative and absolute abundances, LFQ was performed using MaxQuant (version 1.6.1.0) (29). For protein relative abundance, 36 Agilent .d files collected from MS analysis for nine lines and four bioreplicated per line were upload to MaxQuant and searched against the wheat protein database (International Wheat Genome Sequencing Consortium [IWGSC], http://www.wheatgenome.org/, version 1.0, 137,029 sequences) with reversed decoy sequences automatically attached by MaxQuant (30). The search was conducted under general LFQ model with 20 ppm for mass tolerance and 1% false discovery rate. Detailed MaxQuant search parameters are listed in supplemental Table S1*B*. The LFQ intensity of each protein was used as protein relative abundance for downstream data analysis. Protein groups were considered if they were identified in greater than two replicates with at least six independently quantified peptides from the four bioreplicates. To further improve data quality, relative SD by mean was also used as an additional filter. To impose this filter, only those protein groups with a relative SD by mean ≤30% or SD ≤30% of overall mean abundance were retained. Approximately 15% of data from each genotype was trimmed in this process (supplemental Fig. S1). To detect the presence of 1Ay21* HWM-GS in integrated lines and to estimate its absolute abundance in comparison with the other five HMW-GSs, an additional iBAQ search (supplemental Table S1*B*) of each cultivar was conducted against dedicated six HMW-GS sequence database collected from UniProt (https://www.uniprot.org/). The detailed HMW-GS composition for each cultivar is summarized in supplemental Table S2*A*.

The fold change in abundance was calculated using the LFQ or iBAQ protein abundance of each integrated line divided by its corresponding parental cultivar, whereas one-way ANOVA with post hoc Tukey's honestly significant difference test was performed for statistical analysis. Proteins with fold change ≥1.2 or ≤0.83 and *p*-value ≤0.05 were defined as DAPs. Proteins that were uniquely present or absent in NILs or parental cultivars were also considered to be DAPs for the purpose of analysis. Common DAPs were defined as those DAPs with the same direction of fold change in both integrated lines compared with a parental cultivar. The functional category assigned to each DAP was that defined by MapMan (31).

Gene Coexpression Analysis

Gene expression data of the common DAPs under nonstress conditions were collated from the Wheat Expression Browser website (http://www.wheat-expression.com/) and used for gene coexpression analysis (32, 33). The first protein ID of each protein group, as the representative protein ID of the group, was used for transcript data collection. The averaged gene expression abundance of A, B, and D genome homologues was calculated and used in the analysis. The degree of gene coexpression between gene pairs was estimated by Spearman's coefficient correlation and conducted on multiorgan, grain-only, leaf-only, spike-only, and root-only datasets. Two of the 115 genes were homologues, resulting in only 114 triads remaining for gene coexpression analysis. Gene pairs were significantly coexpressed if R ≥0.7, whereas gene pairs were considered as not coexpressed if R ≤0.7.

Identification of cis*-Acting Regulatory Elements on Promoter Regions*

The Chinese Spring genome sequence published by IWGSC was used as the wheat reference sequence for this analysis. The 1500-bp upstream promoter sequence of each gene, starting from the translation initiation codon, was downloaded from Ensembl Plants (https://plants.ensembl.org/index.html). In total, promoter sequences of 240 gene homologues representing 107 of the 144 common DAPs encoding genes were collected. Promoter sequences for two background datasets were also collected. Background dataset 1 included 894 genes representing 1827 redundant non-DAPs across the three cultivars, while background dataset 2 contained 964 genes that were randomly sampled from the IWGSC database (from the 137,052 total set of genes, 1000 genes were sampled but 36 genes failed the promoter sequence collection). These promoter sequences were then submitted to PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (34) for *cis*-element identification.

RESULTS

Alteration in HMW-GS Composition and Total GPC

Through a conventional cross-breeding process, an expressed *Glu-Ay* allele from an Italian wheat line was successfully introgressed into three Australian bread wheat cultivars, namely Gregory, Bonnie Rock, and Yitpi. In each cross, *Ax1* or *Ax1** and Ay^{NE} (not expressed) genes were replaced by

Ax21 and *Ay21** genes, respectively (supplemental Table S2*A*) (20). From these crosses, NILs were developed in each cultivar at the BC4F4 generation (backcrossed for four generations followed by self-pollination for three generations from the F1 seeds). In this study, three lines including one parental cultivar and two Ay-integrated NILs for each of the three cultivars were subjected to a quantitative proteome analysis. Total protein extracts from mature grain samples were extracted, digested with trypsin, and analyzed by LC-MS/MS. This yielded large spectral libraries of MS/MS data and comparable datasets of precursor ion intensities that were amenable to LFQ.

Initially, we sought to use these data to confirm the presence of the Ay HMW-GS protein in NILs and measure the abundance of other HMW-GSs in parental cultivars by searching the data against the peptide sequences listed in supplemental Table S2*A*. These data were expressed as iBAQ values that allow comparison between the same protein in different genotypes and also between different protein types within the genotype (35). This confirmed that five HMW-GSs were detected in parental cultivars, and peptides matching to the sixth Ay HMW-GS protein were only detected in significant abundance in NILs (Fig. 1). The Ay HMW-GS amount detected in NILs was approximately 15% of all y-type HMW-GS protein abundance and approximately 6% of total HMW-GS content. There was some background detection of Ay HMW-GS–like peptides in Gregory, but these were only approximately 11% of the peptide abundances of Ay HMW-GS in the Gregory NILs (supplemental Table S2). Apart from the presence of Ay HMW-GS, no significant absolute abundance changes were detected for peptides matching the other five HMW-GSs in any of the NILs compared with their corresponding parental cultivar. This evidence that introgression of Ay HMW-GS did not affect the absolute abundance of the other five HMW-GSs is consistent with previous results based on HPLC analysis of HMW-GSs (20).

The x-type to y-type HMW-GS ratio is critical to the dough quality and bread-making quality, with lower numbers correlating with higher quality (20, 36). Our label-free quantitative analysis showed x-type/y-type ratio of 1.48, 1.41, and 1.11 for Gregory, Bonnie Rock, and Yitpi parental cultivars, respectively (Fig. 1). This ratio marginally decreased in the presence of Ay HMW-GS in NILs; although only one was statistically significant, the decrease was found in Bonnie Rock–integrated line 1 (Fig. 1*B*). We also found that some of the glutenin subunits, including Ay HMW-GSs, Bx HMW-GSs, and Dx HMW-GSs in Bonnie Rock–integrated line 2 and Dy HMW-GS in Yitpi-integrated line 2, significantly decreased in their relative contribution to the whole HMW-GS pool (supplemental Table S2, *B–D*). Overall, this showed that while the introgression of Ay HMW-GS significantly changed the composition of the HMW-GSs pool in some lines, the absolute abundance of all HMW-GSs and x-type to y-type HMW-GS ratio of majority lines did not statistically change.

It has been reported that the introgression of Ay HMW-GS can increase the total protein content and gluten content of the grain (20, 21). The total protein content on a dry-weight basis showed a significant increase in grain protein in five of the six integrated NILs compared with their parental cultivar with the amount of grain protein on a grain-weight basis being approximately 1% (w/w) higher in the NILs (Table 1). The total gluten content on a protein basis increased by about 5% in four of the six NILs. Given the total HMW glutenin content contributes 2 to 3% to the grain weight (37), the increased amount of Ay HMW-GS alone would contribute less than 0.2% (w/w) of the total protein amount increase observed, suggesting that a substantial proportion of the observed increase was not derived from the extra Ay HMW-GS but from other grain proteins increasing in abundance.

Impacts of the Introgression of 1Ay21 HMW-GS on Grain Proteome Profiles*

To find out which other wheat grain proteins changed in abundance and were contributing to the protein content increase (Table 1), we conducted a systematic analysis of the proteomics data from both parental and integrated lines for all three cultivars. There were high correlations between peptide relative abundances measured across the four bioreplicates (supplemental Table S3). The Pearson's correlation coefficient ranged from 0.70 to 0.90 between replicates, with an average value of 0.79. In total, over 11,000 quantifiable peptides were detected across three cultivars, and these mapped to 683, 648, and 722 proteins from EGA Gregory, Bonnie Rock, and Yitpi, respectively (supplemental Table S4). A principal component analysis of protein relative abundance showed that the three cultivars could be separated by a single principal component explaining nearly 62% of the variation, while there were only minor variations observed between NILs of the same cultivar (Fig. 2*A*). The principal component analysis of data from each individual cultivar indicated that 50% to 60% of the variation could be explained by two components and the parental cultivars were clearly separated from the integrated lines (Fig. 2*B*). Further hierarchical clustering analysis of protein relative abundance data showed that the Ay-integrated lines were more similar to each other than to the parental cultivar, for both Bonnie Rock and Gregory. Similarly, replicates of the same line had a closer relationship to each other than to other lines with some exceptions found in Gregory samples (Fig. 2*C*).

Focusing on proteins found in a parental cultivar and that had the same pattern of change in abundance across both of the integrated NILs, we identified 34 DAPs that increased in abundance and 13 DAPs that decreased in abundance in Gregory (Fig. 3*A*). For Bonnie Rock and Yitpi, these numbers were 14 and 16, and 6 and 10, respectively (Fig. 3, *B* and *C*). While ribosome protein subunits were found as DAPs both increasing and decreasing in abundance, storage proteins typically increased in abundance, whereas biotic and abiotic

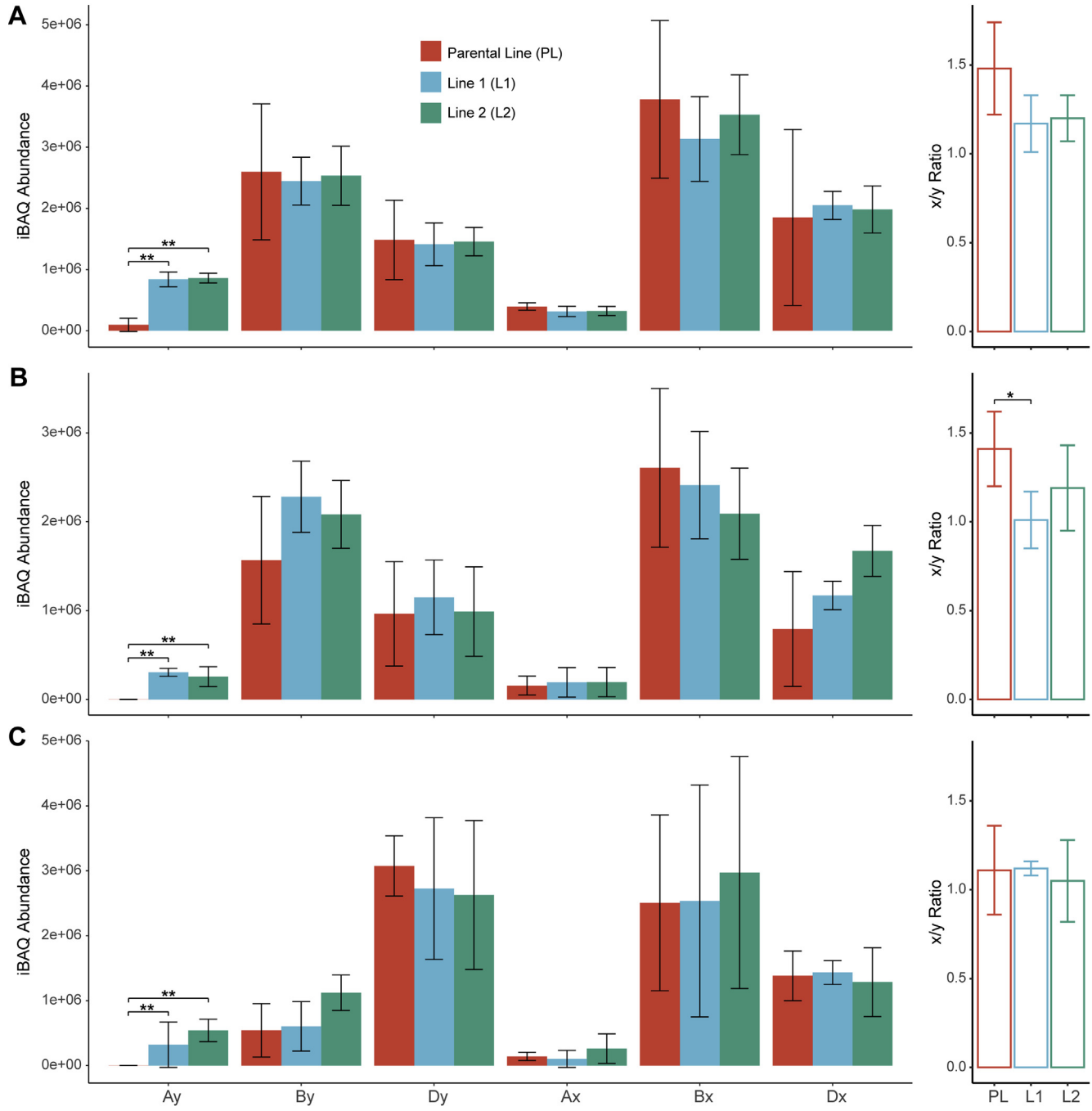


FIG. 1. **HMW-GS abundance in parental and Ay-integrated lines.** HMW-GS abundance estimated through intensity-based absolute quantification (iBAQ) analysis (MaxQuant) for four biological replicates. Detailed results are shown in supplemental Table S2. The *hollow bars* on the *right panel* represent the ratio of x-type and y-type HMW-GSs for each line. Error bars represent the SD. *A*, HMW-GS absolute abundance for cv. Gregory lines. *B*, HMW-GS absolute abundance for cv. Bonnie Rock lines. *C*, HMW-GS absolute abundance for cv. Yitpi lines. Significant abundance changes between lines are marked by *asterisks*, * ≤0.05; ** ≤0.01. HMW-GS, high-molecular-weight glutenin subunit.

stress response proteins typically decreased in abundance. This approach to analysis, however, filtered out present/absent differences that could only be uniquely detected in either parental cultivar or the NILs. Including such present/absent proteins (Fig. 3*D*) more than doubled the number of DAPs that could be considered in the analysis (supplemental Fig. S2; supplemental Table S4, *H* and *I*).

To investigate the biological functions of this expanded set of DAPs (Fig. 3, *A–D*), functional categories of all proteins increasing in abundance (Fig. 3*E*) and decreasing in

TABLE 1
Total wheat grain protein and gluten content in Ay-integrated lines

Cultivar	Line	Total protein content		Gluten content	
		Protein (%)	Net increase (%)	Gluten (%)	Net increase (%)
Gregory	Parent	11.76 ± 0.36	–	22.07 ± 0.87	–
	Line 1	12.40 ± 0.10	0.64*	22.80 ± 0.91	0.73
	Line 2	12.06 ± 0.23	0.30	23.93 ± 0.31	1.86*
Bonnie Rock	Parent	11.77 ± 0.29	–	21.97 ± 0.42	–
	Line 1	12.47 ± 0.15	0.70*	27.30 ± 0.56	5.33**
	Line 2	12.33 ± 0.17	0.56*	27.42 ± 0.26	5.45**
Yitpi	Parent	12.10 ± 0.31	–	25.07 ± 0.62	–
	Line 1	13.43 ± 0.28	1.33**	30.39 ± 0.64	5.32**
	Line 2	13.23 ± 0.11	1.13**	30.03 ± 0.80	4.96**

The averaged percentage protein (w/w) on a grain weight basis and the percentage gluten content (w/w) on a total protein basis with SD (±SD) are shown (n = 3). Significant changes were determined using *t* test; $*p \leq 0.05$, $**p \leq 0.01$.

abundance (Fig. 3*F*) were analyzed. The most frequent DAPs increasing in abundance were storage proteins (33 proteins in total), which were more than three times the number of storage proteins decreasing in abundance. Most of the storage proteins present in higher abundance were identified from protein sets found in parent and NILs (Fig. 3, *A*–*C*), whereas storage proteins with lower abundance were typically identified from the present/absent lists (Fig. 3*D*). Protein synthesis–related proteins and chaperones involved in protein folding and stabilization were similarly abundant among the DAPs increasing and decreasing in abundance. A higher proportion of the DAPs with low abundance were in the stress-response functional category (Fig. 3, *E* and *F*).

As the Ay introgression events in each cultivar were on the long arm of chromosome 1A, we mapped the genome location of the genes encoding DAPs to see if there were any patterns that could have arisen from this genetic location. In total, 159 DAPs from Gregory were successfully mapped, of which, the number on the D subgenome was double that on the A and B subgenome (supplemental Fig. S3*A*). Except for the apparent absence of Bonnie Rock DAPs mapped to chromosome 7B, the same pattern was also observed from the other two cultivars. The higher number of DAPs arising from genes located in the D subgenome was consistent with the higher number of total identified proteins also mapping to the D subgenome, and there were no significant differences between subgenomes in terms of this proportion (supplemental Fig. S3*B*). From this, we concluded there was no gene location pattern of DAPs associated with the chromosome 1A integration site.

To determine which DAPs contributed most to the absolute protein abundance difference between parent and integrated lines, the 20 DAPs with the highest abundance for each parental cultivar were selected based on their absolute abundance estimated by iBAQ scores. These 20 DAPs made up 12 to 17% of the absolute protein abundance detected by MS in each cultivar and consisted mainly of storage proteins, stress-response proteins, and ribosomal proteins (Table 2). With net increase of abundance defined *via* subtracting the iBAQ values of DAPs in integrated lines from its parental cultivar, the total contributions of these 20 abundant DAPs varied from 1% to 5% of total protein abundance. The serpin protein (TraesCS5D01G368900.1), the most abundant DAP in Gregory, contributed most to the total protein abundance increase in both integrated lines, being responsible for a net increase of 0.45% in line 1 and 1.59% in line 2. Likewise, the most abundant two DAPs of Bonnie Rock, alpha-amylase inhibitor (TraesCS4B01G328100.1) and 11S globulin (TraesCS1D01G067100.1), were the biggest contributors in line 1 (1.23%) and line 2 (0.98%), respectively. The heat shock protein (HSP) (TraesCS3A01G033900.1) and globulin 1 (TraesCS5B01G434100.1) of cultivar Yitpi were the top contributors to integrated line 1 (0.96%) and line 2 (0.81%), respectively.

Common Changes in Protein Abundance Induced by the Introgression of 1Ay21 HMW-GS*

Among the redundant set of 503 DAPs across all the cultivars shown in Figure 3, we defined the common DAPs as those that had the same pattern of change in abundance across both integrated lines compared with their parental cultivar. In total, 115 DAPs were selected, including 86 that increased and 29 that decreased in abundance (supplemental Table S4, *H* and *I*). Nearly half of the common DAPs were storage proteins or proteins annotated to be involved in stress response, protein synthesis, protein degradation, or grain development. All the storage proteins in the set of 115 common DAPs showed higher abundance in integrated lines than in their parental cultivar.

Gene Coexpression Profiles of the DAPs Common to Introgression Lines

Expression of a gene that encoded a particular protein during grain development is likely to be a significant factor in determining the final abundance of a protein product. To determine if there was coexpression of the genes encoding common DAPs, we investigated their RNA expression profiles

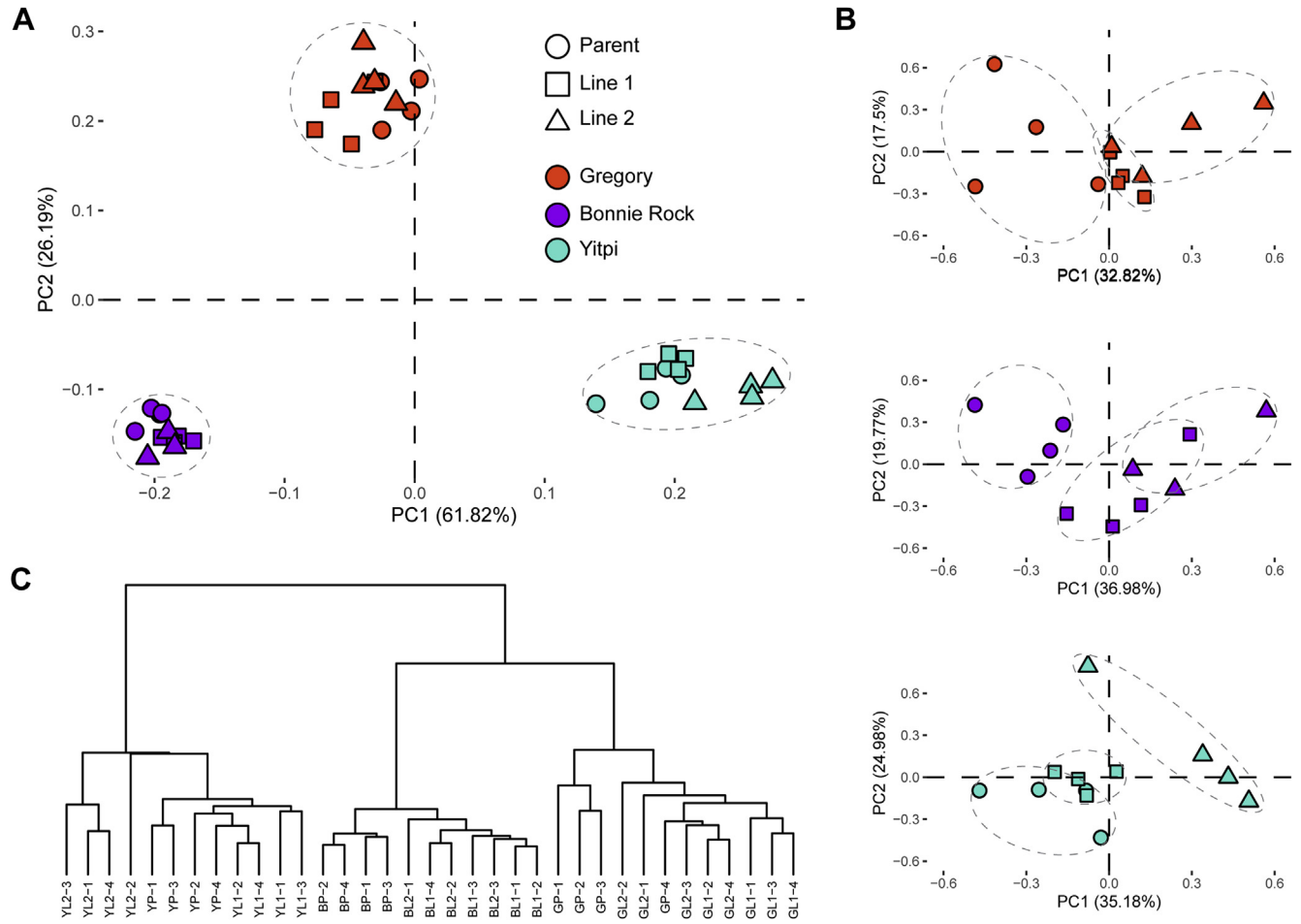


FIG. 2. **Principal component analysis of protein abundances from parental and Ay-integrated lines.** *A*, the PCA for all 35 samples from three cultivars. *B*, the individual PCA for each cultivar. *C*, the dendrogram of hierarchical clustering analysis for the same data used in panel *A*. A dataset of 185 proteins that have relative protein abundance value across all 35 samples was used in panel *A*, while the dataset in panel *B* included 353 proteins for cv. Gregory, 404 for cv. Bonnie Rock, and 432 for cv. Yitpi. The full data used in this analysis are provided in supplemental Table S4, *D–G*. The *dashed circles* outline biological replicates for each cultivar in panel *A* and each line in panel *B*. YP and YL represent the cv. Yitpi parental and integrated line, while G and B represent cv. Gregory and cv. Bonnie Rock, respectively. PCA, principal component analysis.

across multiple wheat varieties, organs, and growth stages. Owing to the fact that wheat genes have three homologues on A, B, and D subgenomes, respectively, and that homologues from the D subgenome show significantly higher expression levels than those from B and A subgenomes (33), the averaged gene expression abundance of all three homologues was used for this analysis. Transcript data for 292 homologous genes representing the common DAPs from 177 samples (23 wheat varieties, 21 organs, and 37 growth stages; Wheat Expression Browser: http://www.wheat-expression.com/) were downloaded and collated and Spearman's coefficient correlation analysis was performed (supplemental Table S5). Hierarchical clustering analysis revealed two distinct gene clusters; a strong coexpression was present within one cluster, whereas weak coexpression was found in the other (Fig. 4*A*). In terms of functional categories, 46 genes for storage proteins, late embryogenesis abundant (LEA) proteins, and HSPs (cluster 1) were strongly coexpressed, whereas 68 genes encoding proteins involved in protein synthesis and degradation were not coexpressed. One exception was a carboxypeptidase that was present in cluster 1 (Fig. 4*A*). Coexpression profiles of the genes for the 114 common DAPs were also conducted for single organs, including the grain, leaf, spike, and root. In comparison with the multiorgan analysis, in the grain, genes encoding storage proteins showed strong coexpression with each other but not with LEA and HSP genes, whereas genes for protein synthesis and degradation-relevant proteins showed similar coexpression profiles (Fig. 4*B*). A similar gene expression profile was observed in the leaf compared with the grain, whereas less significant coexpression was found between genes of the 114 common DAPs in the spike and root (supplemental Fig. S4).

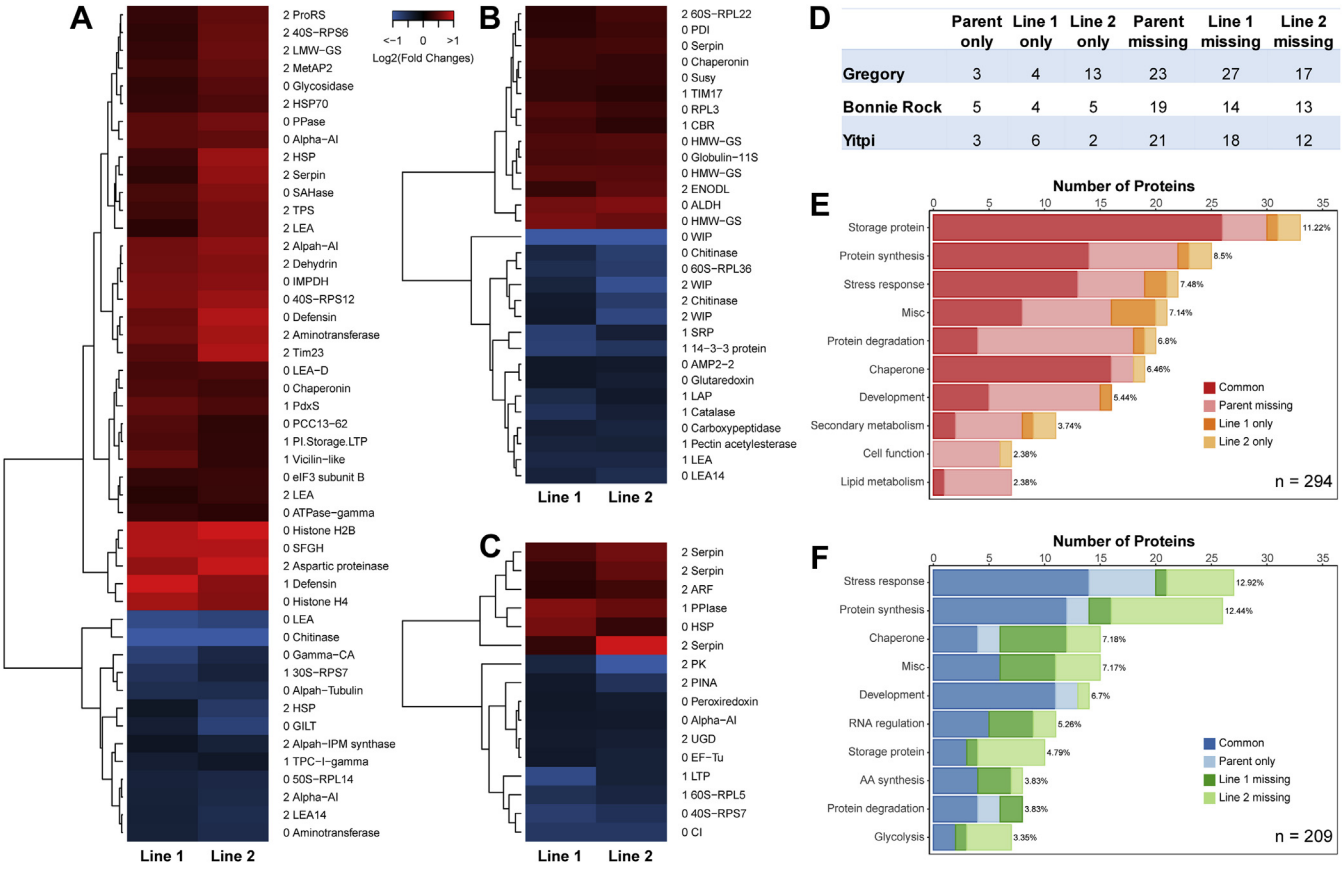


FIG. 3. **Differentially accumulated proteins after integration of Ay HMW-GS across three wheat cultivars.** The relative abundance for each protein was calculated through MaxQuant LFQ analysis. Fold changes in abundance for each protein were calculated by dividing the integrated line value by the parental cultivar value. Only uniquely detected proteins and commonly detected proteins that had fold change ≥1.2-fold or ≤0.83-fold and one-way ANOVA *p*-value ≤0.05 were defined as differentially accumulated proteins (DAPs). The full data are shown in supplemental Table S4. Heat maps only showed DAPs that have the same fold-change pattern (increase or decrease) in both integrated lines. The abbreviation of protein names are attached on the right side of heat maps. The first number of the protein label represents *t* test *p*-value ≤0.05 in both lines (0), in line 1 only (1), and in line 2 only (2). *A*, DAPs for cv. Gregory. *B*, DAPs for cv. Bonnie Rock. *C*, DAPs for cv. Yitpi. *D*, the number of uniquely detected proteins for three cultivars. Parent only, line 1 only, and line 2 only represent those proteins were only detectable in parental cultivar integrated line 1, or integrated line 2, respectively. Parent missing, line 1 missing, and line 2 missing represent those proteins were not detectable in one of the three lines. *E*, MapMan functional category analysis for 294 redundant DAPs that increased in abundance were collected from all three cultivars. *F*, MapMan functional category analysis for 209 redundant DAPs that decreased in abundance was collected from all three cultivars. Only the top ten functional categories are displayed. The summed proportion of all four groups for each functional category is shown. HMW-GS, high-molecular-weight glutenin subunit; LFQ, label-free quantification.

Analysis of cis*-Acting Elements Present in Associated Wheat Gene Promoter Regions*

As gene coexpression can indicate cotranscriptional activation by common transcription factors (TFs), common TF-binding sites may be present in upstream promoter regions of the genes of interest in this analysis. To assess this, the *cis*-acting regulatory elements located in the upstream promoter region of 204 gene homologues representing 107 common DAPs were analyzed using PlantCARE (supplemental Table S6, *A–C*). In total, 140 *cis*-elements belonging to seven subcategories were identified. Most of these *cis*-elements are associated with physiological processes including light response, developmental regulation, hormonal response, and environmental stress response. A total of 894 genes that encode 1827 redundant non-DAPs identified in grain proteomes were used as the control. Statistical analysis suggested that promoter-related elements (TATA-box and CAAT-box) and the majority of light-responsive (G-box, G-Box, Box-4, GATT-motif, GT1-motif, and Pc-CMA2c) and hormone-responsive (ABRE, P-box, and AuxRR-core) *cis*-elements were significantly enriched in the DAP set compared with the control non-DAP set (Table 3; supplemental Table S6*D*). *Cis*-elements related to development (circadian and MBS) and environmental stress response (GC motif and Sp1) showed the opposite pattern with higher enrichment in the non-DAP background genes than in common DAP genes. This analysis was repeated using a randomly sampled set of

TABLE 2
The 20 most abundant differentially accumulated proteins (DAPs) in the Ay-integrated lines of each of three wheat cultivars

First accession	Protein name	iBAQ score	Proportion (%)	Line 1		Line 2	
				FC	*p*-value	FC	*p*-value
cv. EGA Gregory							
TraesCS5D01G368900.1	Serpin family protein	1,511,675	5.9	1.08	0.684	1.27	0.036
TraesCS1A01G066100.1	11S globulin seed storage protein	979,705	3.82	0.98	0.974	1.34	0.025
TraesCS5D01G425800.1	Serpin family protein	447442.5	1.75	1.22	0.247	1.7	0.001
TraesCS3B01G111200.1	Dimeric alpha-amylase inhibitor	199942.5	0.78	1.42	0.029	1.46	0.018
TraesCS6B01G383500.1	Dehydrin	178,020	0.69	1.54	0.063	1.62	0.033
TraesCS5B01G012300.1	Enolase	147929.3	0.58	1.2	0.091	1.24	0.046
TraesCS5A01G469300.1	40S ribosomal protein S26	140878.3	0.55	1.17	0.067	1.27	0.007
TraesCS1D01G067100.1	11S globulin seed storage protein	138,015	0.54	0.97	0.948	1.39	0.014
TraesCS3A01G305400.1	Aspartate aminotransferase	108183.8	0.42	1.14	0.102	1.24	0.008
TraesCS5D01G395600.1	Late embryogenesis abundant D-like protein	96,284	0.38	1.34	0.037	1.38	0.022
TraesCS1B01G144300.1	50S ribosomal protein L5	93947.25	0.37	1.3	0.01	1.14	0.216
TraesCS5D01G479200.1	Wound-induced protease inhibitor	62676.5	0.24	0.61	NA	Absent	NA
TraesCS1A01G372700.1	Late embryogenesis abundant protein	54,874	0.21	1.2	0.099	1.27	0.028
TraesCS7D01G313800.1	60S acidic ribosomal protein P1	48198.5	0.19	Absent	NA	0.89	NA
TraesCS6B01G424800.1	50S ribosomal protein L14	42962.25	0.17	0.75	0.016	0.72	0.009
TraesCS7D01G239400.1	Caffeoyl-CoA O-methyltransferase	36,792	0.14	Absent	NA	0.94	NA
TraesCS7B01G067000.1	Aldose 1-epimerase	36,131	0.14	0.65	0.011	0.84	0.248
TraesCS5B01G179800.1	Phosphoenolpyruvate carboxylase	34,152	0.13	0.81	0.04	0.96	0.765
TraesCS2D01G206300.1	Aquaporin	33047.5	0.13	1.06	0.61	1.3	0.003
TraesCS5D01G454800.1	Thioredoxin	23,402	0.09	Absent	NA	1.2	NA
cv. Bonnie Rock							
TraesCS4B01G328100.1	Dimeric alpha-amylase inhibitor	2,569,800	5.97	1.21	0.019	0.98	0.926
TraesCS1D01G067100.1	11S globulin seed storage protein	1,155,700	2.68	1.35	0.002	1.37	0.002
TraesCS1B01G330000.1	High-molecular-weight glutenin subunit	517,415	1.2	1.58	0	1.51	0.001
TraesCS4B01G262700.1	Vicilin-like antimicrobial peptides 2-2	371162.5	0.86	0.82	0.024	0.8	0.021
TraesCS7D01G133500.1	Alkaline alpha-galactosidase seed imbibition protein	330,510	0.77	1.17	NA	Absent	NA
TraesCS4A01G214200.1	Protein disulfide isomerase	329817.5	0.77	1.22	0	1.29	0
TraesCS5D01G425800.1	Serpin family protein	223,240	0.52	1.23	0.012	1.18	0.062
TraesCS2B01G381600.1	Glutaredoxin	152,480	0.35	0.82	0.008	0.78	0.004
TraesCS4D01G046400.2	14-3-3 protein	129182.5	0.3	0.83	0.019	0.85	0.055
TraesCS7D01G335000.1	60S ribosomal protein L22, putative	105966.3	0.25	1.21	0.114	1.33	0.027
TraesCS1D01G266100.1	Wound-induced protease inhibitor	85674.67	0.2	0.81	0.206	0.57	0.012
TraesCS4D01G109200.1	Ribosomal protein L3	77344.25	0.18	1.38	0	1.28	0.003
TraesCS5A01G385600.1	Late embryogenesis abundant D-like protein	70,333	0.16	1.1	0.195	0.79	0.01
TraesCS6A01G093200.1	Endoglucanase	68401.25	0.16	1.21	0.048	1.2	0.085
TraesCS6A01G049400.1	Alpha-gliadin	50,441	0.12	Absent	NA	1.31	NA
TraesCS2B01G004800.1	Alpha-amylase inhibitor protein	39,659	0.09	Absent	NA	0.51	NA
TraesCS5D01G150000.2	ADP-ribosylation factor	35927.25	0.08	1.12	0.494	1.33	0.04
TraesCS5B01G478200.1	Wound-induced protease inhibitor	32010.5	0.07	0.74	0.235	0.54	0.04
TraesCS7D01G340400.1	Late embryogenesis abundant D-like protein	31,912	0.07	0.96	0.534	0.73	0.001
TraesCS1D01G144000.2	60 kDa chaperonin	31585.75	0.07	1.3	0.01	1.26	0.029
cv. Yitpi							
TraesCS5D01G368900.1	Serpin family protein	1,485,900	2.05	1.08	0.467	1.29	0.004
TraesCS5B01G434100.1	Globulin-1	1,450,825	2	1.1	0.517	1.4	0.003
TraesCS3A01G033900.1	Heat shock protein	1,261,755	1.74	1.55	0	1.26	0.021
TraesCS5B01G488800.1	rRNA N-glycosidase	1,144,263	1.58	1.02	0.965	1.3	0.022
TraesCS1D01G317300.1	High-molecular-weight glutenin subunit	1,069,063	1.47	1.11	0.628	1.55	0.002
TraesCS7D01G066400.1	Peroxidase	539,750	0.74	1.05	0.798	1.26	0.015
TraesCS5A01G417800.1	Serpin family protein	392782.5	0.54	1.35	0.073	1.53	0.009
TraesCS1B01G034400.1	Chymotrypsin inhibitor	278293.3	0.38	0.63	0.03	0.63	0.029
TraesCS1B01G081000.1	Protease inhibitor/seed storage/lipid transfer protein family protein	195,055	0.27	0.93	0.55	0.75	0.012

TABLE 2—*Continued*

First accession	Protein name	iBAQ score	Proportion (%)	Line 1		Line 2	
				FC	*p*-value	FC	*p*-value
TraesCS2D01G261500.1	Eukaryotic translation initiation factor 5A	109,729	0.15	1.31	0.033	1.18	0.217
TraesCS2B01G047400.1	Heat shock protein 90	109451.8	0.15	1.25	0.01	1.03	0.868
TraesCS6D01G049100.1	Heat shock 70-kDa protein	80967.75	0.11	0.8	0.002	0.85	0.014
TraesCS2D01G128900.1	Cysteine proteinase inhibitor	72805.75	0.1	0.59	0.003	0.96	0.899
TraesCS2D01G348800.1	Chitinase	35795.5	0.05	0.57	0.016	1.14	0.522
TraesCS7D01G411700.1	Defensin	32,103	0.04	0.4	NA	Absent	NA
TraesCS7B01G184100.1	HSP20-like chaperones superfamily protein isoform 3	31421.5	0.04	0.87	0.315	1.27	0.043
TraesCS1D01G424500.1	Programmed cell death protein 4	29,581	0.04	1.25	0.036	1.17	0.105
TraesCS5B01G003500.1	Grain softness protein	25,665	0.04	Absent	NA	0.32	NA
TraesCS1D01G306800.1	60S ribosomal protein L30	25,076	0.03	Absent	NA	0.86	NA
TraesCS6A01G371100.1	Aldehyde dehydrogenase	24480.5	0.03	1.17	0.006	0.96	0.611

The proportion column contains the percentage of iBAQ score of each protein to the summed protein iBAQ score. The fold change (FC) and *t* test *p*-value were collected from supplemental Table S4.

1000 wheat genes as a background dataset and that returned a result highly consistent with using the 894 non-DAP gene set as the control (supplemental Table S6*E*). Comparison of the two background datasets showed no statistical differences in *cis*-element enrichment between them (supplemental Table S6*F*).

DISCUSSION

Introgression of Ay genes has been trialed for GPC improvement in wheat breeding programs worldwide (3, 7, 8, 17). Many studies have been conducted to either characterize Ay gene structures or explore its economic potential in modern wheat breeding programs (16). These reports have mainly focused on gene discovery and gene structure characterization (10, 23, 38–40) and the influence of Ay HMW-GS introgression on agronomic and quality traits, including grain yield, GPC, storage protein composition and dough quality (17, 20, 21). By studying the proteome of Ay-integrated NILs, we have provided insights into the nature of the changes in the wider grain protein profile of bread wheat induced by Ay introgression events and propose a model to explain the molecular events observed.

Introgression of the Expressed Ay *Gene Changes More Than the Abundance of* Ay *Itself*

The initial intent of cross-breeding the *Glu-1Ay* gene into hexaploid wheat was to improve the GPC and quality by addition of an extra Ay HMW-GS. Previous work has proved that the Ay HMW-GS–integrated lines have a higher GPC (17, 20). However, it was unknown whether the increased protein abundance was solely from the Ay HMW-GS protein itself or if there were other wheat grain proteins contributing to the protein content improvement. Our results showed that the integrated lines had a significantly higher protein content than parental cultivars as well as an increase in the gluten content (Table 1). Calculations based on the contribution of HMW-GS to total gluten (37) showed that the total increase in all gluten proteins represented on average only 52% of the total protein content increase. The proteome data showed that approximately a quarter of the proteins identified in wheat grains by MS significantly changed in their abundance in Ay-integrated lines (Fig. 3). The most abundant 20 DAPs of each cultivar (predominantly storage proteins, stress-response proteins, and ribosomal proteins) contributed most of the remaining 48% increase of total protein abundance (Table 2). The functional category analysis showed these common DAPs belonged to 36 different categories, indicating it is not a single system or a discrete part of the cellular machinery that is altered after Ay introgression. The genomic locus analysis showed that DAPs arise from genes that are relatively evenly distributed across the 21 chromosomes of bread wheat, so changes in gene expression that might be driving these changes in protein abundance are not tightly associated with the introgression event on chromosome 1A. It appears expression of Ay HMW-GS has a significant downstream impact resulting in alterations to the wheat grain proteome, which causes the significant increase in grain protein. The *cis*-element analysis conducted provides support for Ay HMW-GS integration-dependent activation of genes encoding common DAPs because their promoters were enriched with TATA box and light-response elements (Table 3). Specific downstream targets involved in ATP synthesis, amino acid synthesis and degradation, protein folding, and RNA regulation (supplemental Table S4, *H* and *I*) can be treated as valuable protein candidates for future breeding programs of the GPC and bread-making quality improvement. However, it was the changes in abundance of three major groups of proteins that dominated the protein response: storage proteins, proteins of the protein synthesis machinery, and proteins involved in protein stability and degradation systems.

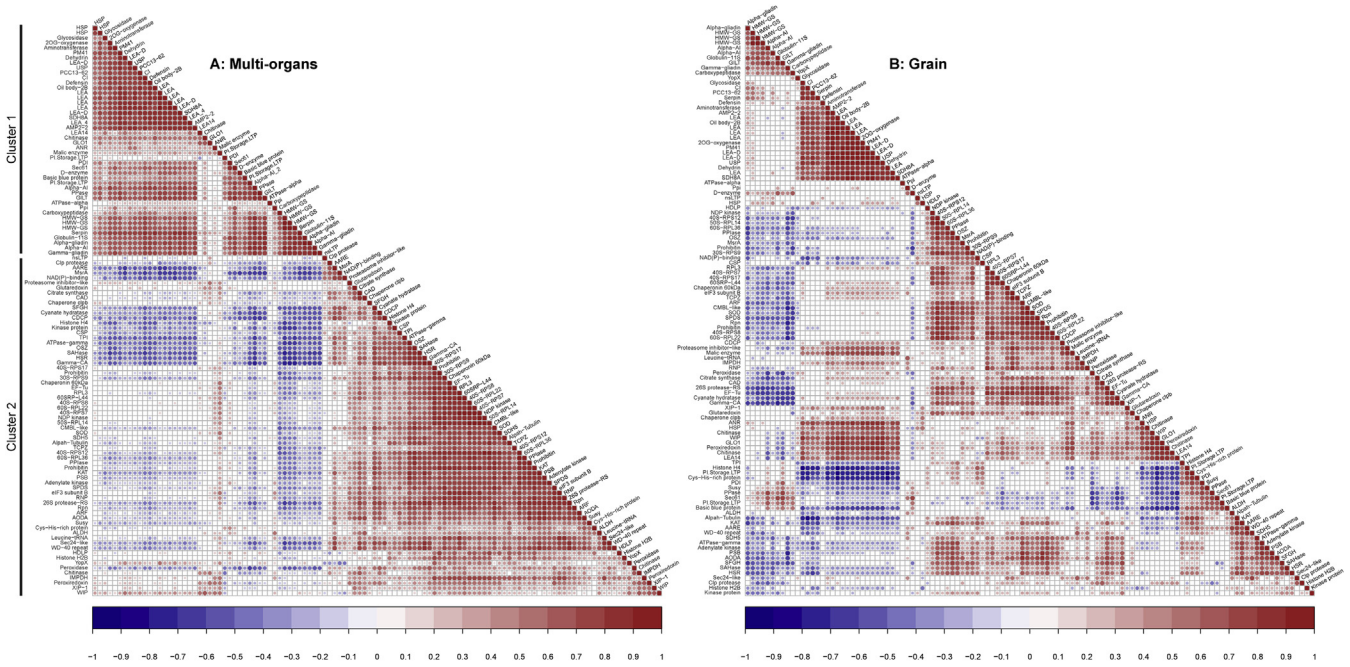


FIG. 4. **Correlation coefficients of gene expression profiles of the 114 common differentially accumulated proteins.** *A*, coefficients of the gene coexpression profiles across 23 varieties, 21 organs, and 37 growth stages (supplemental Table S5). *B*, coefficients of the gene coexpression profiles in grain-only samples. The gene expression data of these general DAPs were collated from the Wheat Expression Browser (33). Spearman's correlation analysis and hierarchical clustering for both columns and rows were used in data visualization. The abbreviated protein names are shown, and protein full names are listed in supplemental Table S5. Pairs with p-value $\geq$0.05 were removed, leaving blank cells on the plot. DAPs, differentially accumulated proteins.

Ay HMW-GS Expression Triggers Changes in the Abundance of Other Storage Proteins

A total of 43 storage proteins, other than Ay HMW-GS, were defined as DAPs in this analysis, among which 33 had higher abundance and ten had lower abundance in Ay-integrated lines. All the HMW-GS (seven proteins), globulin-11S (five proteins), and all but one of the serpins (ten proteins) increased in abundance. These HMW-GSs are key determinants in the dough-quality formation and bread-making quality and may be important contributors to the quality traits attributed to Ay integrated lines (3, 21). Globulin-11S is the main form of the salt-soluble globulin proteins that account for approximately 10% of grain storage protein (37, 41). Although the role of globulin-11S in bread-making quality is not very clear, evidence indicates that it is able to form interchain and intrachain disulfide links and partially contributes to variations found in the dough quality (41–43). Serpins make up 4% of total grain protein and may contribute to the increase of GPC through their role in inhibiting proteases, thereby increasing the stability of storage proteins (44, 45). Serpins can influence the bread-making quality either *via* modifying the molecular structure of prolamin storage proteins or by forming intermolecular disulfide bridges between serpins and between serpins and β-amylase proteins (46). Taken together, the accumulation of these three protein families (especially HMW-GSs) is likely to be the major driving force in the improvement of GPC and bread-making quality observed in Ay-integrated lines. Better bread-making quality in Ay-integrated lines in the Bonnie Rock genetic background has been reported (20), and it was in this background that the largest and most consistent increases in the gluten content and HMW-GSs were observed in this study (Figs. 1 and 3, and Table 1).

Ay HMW-GS Induced Selective Changes in Protein-Synthesis Machinery

As the basic functional units of protein-synthesis machinery, ribosomes and their function throughout grain development will be important for a higher GPC. Eukaryotic ribosomes are composed of small 40S and large 60S subunits containing up to 100 different proteins. Ribosomes are initially assembled in the nucleolus and then released into the cytoplasm as preribosomal particles (47, 48). In plants, the specific isoforms of a given ribosome protein can differ between tissue types (49) and are exchanged during environmental stress such as cold and oxidative stress (50–52). Ribosome composition is critical for the speed and efficiency of protein synthesis, especially of proteins containing rare codons (53). In this study, 43 specific ribosomal proteins showed significant changes in abundance, although the overall net increase in amount was minor within Ay-integrated lines compared with

TABLE 3
Cis-acting regulatory element analysis of upstream 1500-bp promoter region of genes encoding differentially accumulated proteins in Ay-integrated lines

Element name	No. of cluster 1		No. of cluster 2		No. of background		*p*-value (Fisher's exact test)			Function category
	Total	Mean	Total	Mean	Total	Mean	Cluster 1 *versus* background 1	Cluster 2 *versus* background 1	Cluster 1 *versus* cluster 2	
Cluster 1 enriched *cis*-element sites										
TATA-box	1504	15.67	2274	15.79	12,265	13.72	0	0	0.129	Promoter-related elements
WUN-motif	35	0.36	30	0.21	186	0.21	0.006	0.921	0.043	Environmental stress–related elements
G-box	308	3.21	336	2.33	2410	2.7	0.013	0.024	0	Light-responsive elements
G-Box	117	1.22	121	0.84	852	0.95	0.027	0.281	0.012	Light-responsive elements
Box 4	73	0.76	69	0.48	457	0.51	0.004	0.751	0.013	Light-responsive elements
GATT-motif	5	0.05	1	0.01	9	0.01	0.009	1	0.045	Light-responsive elements
ABRE	454	4.73	400	2.78	3109	3.48	0	0	0	Hormone-responsive elements
P-box	37	0.39	37	0.26	225	0.25	0.029	0.857	0.124	Hormone-responsive elements
Cluster 2 enriched *cis*-element sites										
CAAT-box	1823	18.99	2888	20.06	16,178	18.1	0.263	0	0	Promoter-related elements
AT-rich element	7	0.07	16	0.11	55	0.06	0.67	0.037	0.397	Site binding–related elements
GT1-motif	45	0.47	91	0.63	445	0.5	0.649	0.032	0.066	Light-responsive elements
Pc-CMA2c	1	0.01	5	0.03	6	0.01	0.518	0.011	0.411	Light-responsive elements
AuxRR-core	10	0.1	27	0.19	104	0.12	0.875	0.029	0.095	Hormone-responsive elements
Non-DAPs enriched *cis*-element sites										
Motif I	9	0.09	1	0.01	46	0.05	0.113	0.017	0.002	Development-related elements
Circadian	6	0.06	4	0.03	111	0.12	0.088	0.001	0.335	Development-related elements
MBS	72	0.75	68	0.47	579	0.65	0.323	0.018	0.012	Environmental stress–related elements
GC motif	58	0.6	75	0.52	673	0.75	0.081	0.003	0.537	Environmental stress–related elements
Sp1	70	0.73	105	0.73	1107	1.24	0	0	0.817	Light-responsive elements
3-AF1 binding site	0	0	4	0.03	34	0.04	0.045	0.813	0.15	Light-responsive elements
A-box	63	0.66	113	0.78	875	0.98	0.001	0.041	0.167	Hormone-responsive elements

Fisher's exact test was applied for statistical analysis, and only those known functions and significantly changed sites are included in this table. The detailed *cis*-element information of each gene and full summary table are listed in supplemental Table S6.

parental cultivars (supplemental Table S4, *H* and *I*). Non-ribosomal proteins that support translation, such as the eukaryotic translation initiation factor 3B and 5A and elongation factor 1 alpha, were DAPs that increased in abundance (supplemental Table S4, *H* and *I*). These complicated patterns might imply that protein synthesis machinery regulation is initiated in Ay lines that may support changes in storage-protein synthesis.

Ay HMW-GS Induced Changes to the Protein Folding, Stability, and Proteolysis Systems in Grains

There is no doubt that keeping newly assembled proteins stable can be as important as making new proteins to achieve a higher protein content (54, 55). Summing up the total abundance increases and decreases for 34 molecular chaperones showed a net accumulation of 2.3-fold. During the grain-maturation period, protein aggregation is triggered by desiccation; hence, having more molecular chaperone proteins will better prevent irreversible protein aggregation and help wheat grain withstand severe desiccation (56). The accumulation of LEA proteins usually occurs during mid to late embryogenesis and correlates with the acquisition of seed desiccation tolerance (57). Summing up the net impact of 30 changes to LEA protein abundance (16 increases and 14 decreases) showed a net increase of 1.7-fold. The combination of accumulation of chaperones and LEA in Ay-integrated lines may benefit a higher GPC through improving the stability of storage proteins during grain development. To maintain protein homeostasis, proteolysis is also required to remove damaged, misfolded, and dysfunctional proteins. The ubiquitin–proteasome system is the principal mechanism responsible for degrading short-lived regulatory proteins and soluble misfolded proteins (58, 59). Autophagy, by contrast,

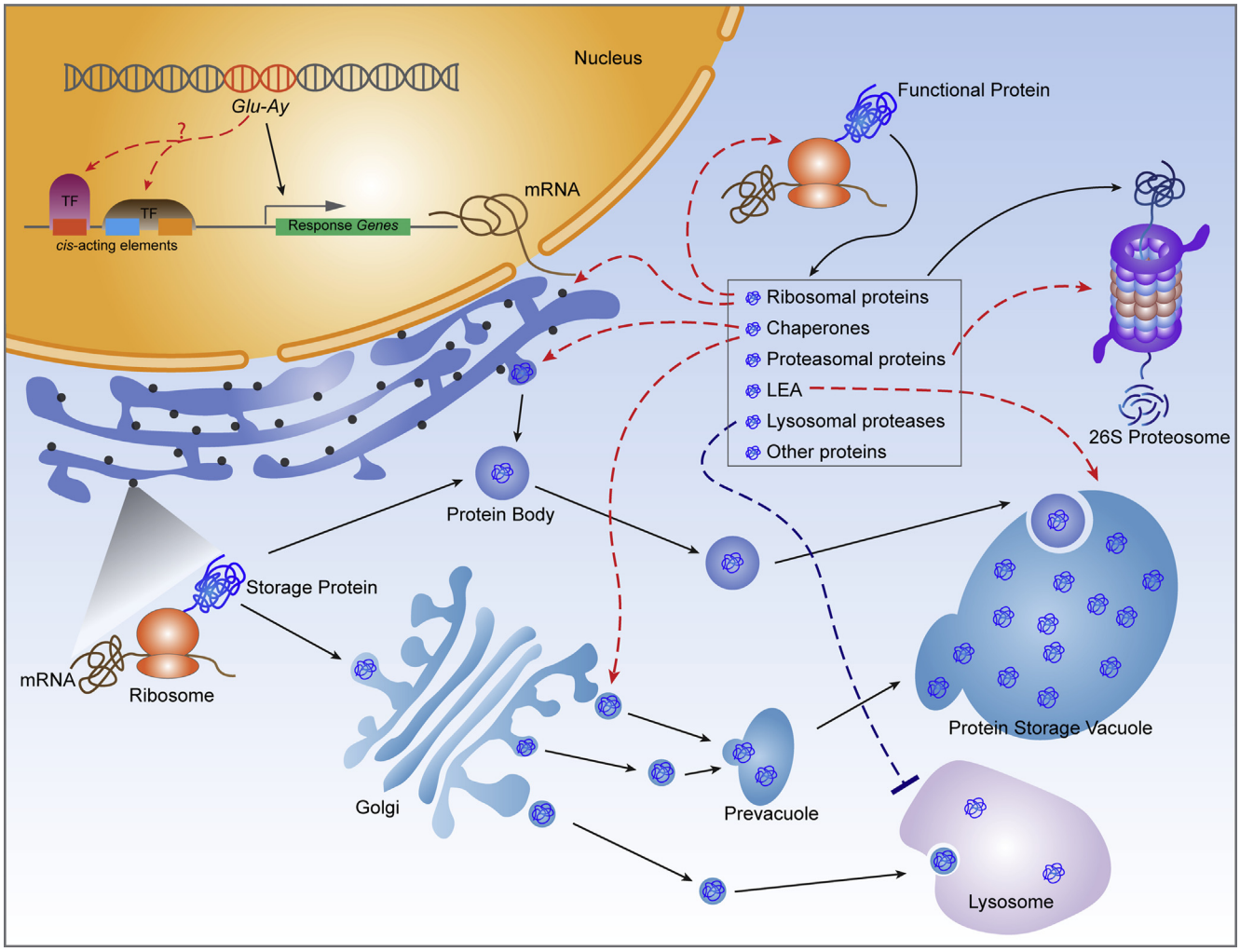


FIG. 5. **Model of a putative pathway for grain protein accumulation triggered by the introgression of 1Ay21* HMW-GS.** *Black arrows* represent general protein trafficking pathway; *red dashed arrows* represent the positive regulation effects, and the *blue dashed flat-headed arrow* represents negative regulation effects. ER, endoplasmic reticulum; HMW-GS, high-molecular-weight glutenin subunit; LEA, late embryogenesis abundant; TF, transcription factor.

can eliminate larger protein complexes and insoluble protein aggregates, and it has been reported to be involved in the degradation of storage proteins during seed germination (60–62). Our results indicated that proteasomal proteins, such as the 26S protease regulatory subunit and proteasome β subunit, showed an increasing pattern in abundance, whereas most lysosomal proteases such as carboxypeptidase and leucine aminopeptidase decreased in abundance in Ay-integrated lines (supplemental Table S4, *H* and *I*). Thus, we could hypothesize that the autophagy–lysosome system was suppressed to stabilize newly synthesized storage proteins during the accumulation of storage protein in Ay lines. In contrast, increased 26S proteasome system could recycle nutrients from regulatory or incorrectly folded proteins to optimize protein abundance.

A Putative Model of Glu-Ay *Induction of High Grain Content and Bread-Making Quality*

Combining our results and previous reports, we proposed a model that could now be tested of how introgression of the *Glu-1Ay21** allele improves the wheat GPC and bread-making quality (Fig. 5). This proposal begins with the expression of the extra Ay gene, taking the total number of HMW-GSs from 5 to 6. The increased HMW-GS content is deposited into protein storage vacuoles through either the Golgi apparatus into the vacuole or the protein body accumulating directly within the lumen of the ER (63, 64). By unknown mechanisms, this Ay-dependent protein body accumulation activates TFs that bind sites in the promoter regions of a suite of genes that already form a developmental gene expression program to

enhance storage proteins and protein synthesis regulatory proteins in wheat grains. Ribosomal composition changes may then fulfill the higher demand for protein synthesis, and more molecular chaperones assist nascent proteins to fold properly and be stored. Owing to the protection from higher abundances of chaperones and LEA, other storage proteins are deposited into protein storage vacuoles with enhanced stability, which led to less need of the lysosome system and a broad increase in the abundance of HMW-GSs, globulin-11S, and serpins. Boosting of the ubiquitin–proteasome system recycles nutrients from unwanted proteins, providing a positive reinforcement and resources for more storage protein synthesis and deposition. As a consequence of these concerted actions, the GPC for Ay HMW-GS–introgressed lines increases, leading to improvement of the dough quality and bread-making quality because of the alteration of gluten composition. Such a model defines a series of steps that would have promise for further enhancement of the protein content in wheat grain. These steps could include the identification of TFs to further enhance expression of the suite of genes identified here to contribute to enhanced protein content, modifying the localization of the Ay protein, and/or changes in the abundance of chaperones or ribosome subunit DAPs identified here to determine if they are causes of consequences of enhanced protein abundance in wheat grains.

DATA AVAILABILITY

The MS raw data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the PRIDE (65) partner repository with dataset identifier PXD021706. Data of the spectra have been deposited through MS-Viewer, and the unique search keys for each sample are listed in supplemental Table S1*C*.

Supplemental data—This article contains supplemental data.

Author contributions—H. C. conceptualization, methodology, software, formal analysis, investigation, visualization, and writing–original draft. O. D. methodology, supervision, and writing–review and editing. S. I. conceptualization, validation, data curation, and writing–review and editing. J. Z. resources and investigation. W. M. supervision, project administration, and funding acquisition. A. H. M. conceptualization, writing–review and editing, supervision, project administration, and funding acquisition.

Funding and additional information—H. C. was supported by Research Training Program Fee-Offset–International Student and UWA Safety-Net Top-Up Scholarship. This work was support by Australian Research Council funding to A. H. M. (CE140100008) and Grains Research and Development Corporation funding to W. M. (UMU00036).

Conflict of interest—The authors declare no competing interests.

Abbreviations—The abbreviations used are: DAP, differentially accumulated proteins; GPC, grain protein content; HMW-GS, high-molecular-weight glutenin subunit; HSP, heat shock proteins; iBAQ, intensity-based absolute quantification; LEA, late embryogenesis abundant; LFQ, label-free quantification; NIL, near-isogenic lines; TF, transcription factors.

Received February 5, 2021, and in revised from, April 15, 2021
Published, MCPRO Papers in Press, May 14, 2021, https://doi.org/10.1016/j.mcpro.2021.100097

REFERENCES

1. Shewry, P. R. (2009) Wheat. *J. Exp. Bot.* **60**, 1537–1553
2. Rasheed, A., Xia, X. C., Yan, Y. M., Appels, R., Mahmood, T., and He, Z. H. (2014) Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications. *J. Cereal Sci.* **60**, 11–24
3. Shewry, P. R., and Halford, N. G. (2002) Cereal seed storage proteins: Structures, properties and role in grain utilization. *J. Exp. Bot.* **53**, 947–958
4. Shewry, P. R., Halford, N. G., Belton, P. S., and Tatham, A. S. (2002) The structure and properties of gluten: An elastic protein from wheat grain. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357**, 133–142
5. Shewry, P. R., Halford, N. G., and Tatham, A. S. (1992) High-molecular-weight subunits of wheat glutenin. *J. Cereal Sci.* **15**, 105–120
6. Nakamura, H. (1999) Identification of alleles for complex gene loci Glu-A1, Glu-B1, and Glu-D1, which code for nigh molecular weight subunits of glutenin in Japanese hexaploid wheat varieties. *J. Agric. Food Chem.* **47**, 5273–5277
7. Payne, P. I., and Lawrence, G. J. (1983) Catalogue of alleles for the complex gene loci, Glu-A1, Glu-B1, and Glu-D1 which code for high-molecular-weight subunits of glutenin in hexaploid wheat. *Cereal Res. Commun.* **11**, 29–35
8. Waines, J. G., and Payne, P. I. (1987) Electrophoretic analysis of the high-molecular-weight glutenin subunits of Triticum monococcum, T. urartu, and the A genome of bread wheat (T. aestivum). *Theor. Appl. Genet.* **74**, 71–76
9. Ciaffi, M., Lafiandra, D., Porceddu, E., and Benedettelli, S. (1993) Storage-protein variation in wild emmer wheat (Triticum turgidum ssp.dicoccoides) from Jordan and Turkey. I. Electrophoretic characterization of genotypes. *Theor. Appl. Genet.* **86**, 474–480
10. Jiang, Q. T., Wei, Y. M., Wang, F., Wang, J. R., Yan, Z. H., and Zheng, Y. L. (2009) Characterization and comparative analysis of HMW glutenin 1Ay alleles with differential expressions. *BMC Plant Biol.* **9**, 16
11. Kiszonas, A. M., and Morris, C. F. (2018) Wheat breeding for quality: A historical review. *Cereal Chem.* **95**, 17–34
12. Shewry, P. R. (2007) Improving the protein content and composition of cereal grain. *J. Cereal Sci.* **46**, 239–250
13. Branlard, G., and Dardevet, M. (1985) Diversity of grain proteins and bread wheat quality. 1. Correlation between gliadin bands and flour quality characteristics. *J. Cereal Sci.* **3**, 329–343
14. Liu, S. W., Zhao, S. Y., Chen, F., and Xia, G. M. (2007) Generation of novel high quality HMW-GS genes in two introgression lines of Triticum aestivum/Agropyron elongatum. *BMC Evol. Biol.* **7**, 76
15. Payne, P. I., Nightingale, M. A., Krattiger, A. F., and Holt, L. M. (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat-varieties. *J. Sci. Food Agric.* **40**, 51–65
16. Margiotta, B., Urbano, M., Colaprico, G., Johansson, E., Buonocore, F., DOvidio, R., and Lafiandra, D. (1996) Detection of y-type subunit at the Glu-A1 locus in some Swedish bread wheat lines. *J. Cereal Sci.* **23**, 203–211
17. Rogers, W. J., Miller, T. E., Payne, P. I., Seekings, J. A., Sayers, E. J., Holt, L. M., and Law, C. N. (1997) Introduction to bread wheat (Triticum aestivum L) and assessment for bread-making quality of alleles from T-boeoticum Boiss ssp thaoudar at Glu-A1 encoding two high-molecular-weight subunits of glutenin. *Euphytica* **93**, 19–29
18. Li, Q. Y., Yan, Y. M., Wang, A. L., An, X. L., Zhang, Y. Z., Hsam, S. L. K., and Zeller, F. J. (2006) Detection of HMW glutenin subunit variations among 205 cultivated emmer accessions (Triticum turgidum ssp dicoccum). *Plant Breed.* **125**, 120–124

19. Xu, L. L., Li, W., Wei, Y. M., and Zheng, Y. L. (2009) Genetic diversity of HMW glutenin subunits in diploid, tetraploid and hexaploid Triticum species. *Genet. Resour. Crop. Evol.* **56**, 377–391
20. Roy, N., Islam, S., Ma, J. H., Lu, M. Q., Torok, K., Tomoskozi, S., Bekes, F., Lafiandra, D., Appels, R., and Ma, W. J. (2018) Expressed Ay HMW glutenin subunit in Australian wheat cultivars indicates a positive effect on wheat quality. *J. Cereal Sci.* **79**, 494–500
21. Roy, N., Islam, S., Yu, Z. T., Lu, M. Q., Lafiandra, D., Zhao, Y., Anwar, M., Mayer, J. E., and Ma, W. J. (2020) Introgression of an expressed HMW 1Ay glutenin subunit allele into bread wheat cv. Lincoln increases grain protein content and breadmaking quality without yield penalty. *Theor. Appl. Genet.* **133**, 517–528
22. Iqbal, M., Navabi, A., Salmon, D. F., Yang, R. C., and Spaner, D. (2007) Simultaneous selection for early maturity, increased grain yield and elevated grain protein content in spring wheat. *Plant Breed.* **126**, 244–250
23. Yu, Z. T., Peng, Y. C., Islam, M. S., She, M. Y., Lu, M. Q., Lafiandra, D., Roy, N., Juhasz, A., Yan, G. J., and Ma, W. J. (2019) Molecular characterization and phylogenetic analysis of active y-type high molecular weight glutenin subunit genes at Glu-A1 locus in wheat. *J. Cereal Sci.* **86**, 9–14
24. Vensel, W. H., Tanaka, C. K., and Altenbach, S. B. (2014) Protein composition of wheat gluten polymer fractions determined by quantitative two-dimensional gel electrophoresis and tandem mass spectrometry. *Proteome Sci.* **12**, 8
25. Wessel, D., and Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute-solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143
26. Schaffner, W., and Weissmann, C. (1973) A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56**, 502–514
27. Yang, F., Shen, Y. F., Camp, D. G., and Smith, R. D. (2012) High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. *Expert Rev. Proteomics* **9**, 129–134
28. Duncan, O., Trosch, J., Fenske, R., Taylor, N. L., and Millar, A. H. (2017) Resource: Mapping the Triticum aestivum proteome. *Plant J.* **89**, 601–616
29. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372
30. Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C. J., Stein, N., Choulet, F., Distelfeld, A., Eversole, K., Poland, J., Rogers, J., Ronen, G., Sharpe, A. G., *et al*. (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, 661
31. Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L. A., Rhee, S. Y., and Stitt, M. (2004) Mapman: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **37**, 914–939
32. Borrill, P., Ramirez-Gonzalez, R., and Uauy, C. (2016) expVIP: A customizable RNA-seq data analysis and visualization platform. *Plant Physiol.* **170**, 2172–2186
33. Ramirez-Gonzalez, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., Khedikar, Y., Robinson, S. J., Cory, A. T., Florio, T., Concia, L., *et al*. (2018) The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089
34. Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P., and Rombauts, S. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**, 325–327
35. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319
36. Leon, E., Piston, F., Aouni, R., Shewry, P. R., Rosell, C. M., Martin, A., and Barro, F. (2010) Pasting properties of transgenic lines of a commercial bread wheat expressing combinations of HMW glutenin subunit genes. *J. Cereal Sci.* **51**, 344–349
37. Zorb, C., Ludewig, U., and Hawkesford, M. J. (2018) Perspective on wheat yield and quality with reduced nitrogen supply. *Trends Plant Sci.* **23**, 1029–1037
38. Bai, J. R., Jia, X., Liu, K. F., and Wang, D. W. (2004) Cloning and characterization of the coding sequences of the 1Ay high molecular weight glutenin subunit genes from Triticum urartu. *Acta Bot. Sin.* **46**, 463–471
39. Bi, Z. G., Wu, B. H., Hu, X. G., Guo, X. H., Liu, D. C., and Zheng, Y. L. (2014) Identification of an active 1Ay gene from Triticum turgidum ssp dicoccoides. *Czech J. Genet. Plant Breed.* **50**, 208–215
40. Gutierrez, M. V., Guzman, C., Martin, L. M., and Alvarez, J. B. (2011) Molecular characterization of the Glu-Ay gene from Triticum urartu for its potential use in quality wheat breeding. *Plant Genet. Resour.* **9**, 334–337
41. Hellemans, T., Landschoot, S., Dewitte, K., Van Bockstaele, F., Vermeir, P., Eeckhout, M., and Haesaert, G. (2018) Impact of crop husbandry practices and environmental conditions on wheat composition and quality: A review. *J. Agric. Food Chem.* **66**, 2491–2509
42. Inquello, V., Raymond, J., and Azanza, J. L. (1993) Disulfide interchange reactions in 11S globulin subunits of Cruciferae seeds. Relationships to gene families. *Eur. J. Biochem.* **217**, 891–895
43. Jung, R., Nam, Y. W., Saalbach, I., Muntz, K., and Nielsen, N. C. (1997) Role of the sulfhydryl redox state and disulfide bonds in processing and assembly of 11S seed globulins. *Plant Cell* **9**, 2037–2050
44. Cane, K., Sharp, P. J., Eagles, H. A., Eastwood, R. F., Hollamby, G. J., Kuchel, H., Lu, M. Q., and Martin, P. J. (2008) The effects on grain quality traits of a grain serpin protein and the VPM1 segment in Southern Australian wheat breeding. *Aust. J. Agric. Res.* **59**, 883–890
45. Ostergaard, H., Rasmussen, S. K., Roberts, T. H., and Hejgaard, J. (2000) Inhibitory serpins from wheat grain with reactive centers resembling glutamine-rich repeats of prolamin storage proteins - cloning and characterization of five major molecular forms. *J. Biol. Chem.* **275**, 33272–33279
46. Roberts, T. H., and Hejgaard, J. (2008) Serpins in plants and green algae. *Funct. Integr. Genomics* **8**, 1–27
47. Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003) Ribosome assembly in eukaryotes. *Gene* **313**, 17–42
48. Panse, V. G., and Johnson, A. W. (2010) Maturation of eukaryotic ribosomes: Acquisition of functionality. *Trends Biochem. Sci.* **35**, 260–266
49. Salih, K. J., Duncan, O., Li, L., Trosch, J., and Millar, A. H. (2020) The composition and turnover of the Arabidopsis thaliana 80S cytosolic ribosome. *Biochem. J.* **477**, 3019–3032
50. Bailey-Serres, J., and Freeling, M. (1990) Hypoxic stress-induced changes in ribosomes of maize seedling roots. *Plant Physiol.* **94**, 1237–1243
51. Byrne, M. E. (2009) A role for the ribosome in development. *Trends Plant Sci.* **14**, 512–519
52. Salih, K. J., Duncan, O., Li, L., O'Leary, B., Fenske, R., Trosch, J., and Millar, A. H. (2020) Impact of oxidative stress on the function, abundance, and turnover of the Arabidopsis 80S cytosolic ribosome. *Plant J.* **103**, 128–139
53. Novoa, E. M., and Ribas de Pouplana, L. (2012) Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet.* **28**, 574–581
54. Fink, A. L. (1999) Chaperone-mediated protein folding. *Physiol. Rev.* **79**, 425–449
55. Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M., and Tsai, F. T. F. (2003) The structure of clpB: A molecular chaperone that rescues proteins from an aggregated state. *Cell* **115**, 229–240
56. Liberek, K., Lewandowska, A., and Zietkiewicz, S. (2008) Chaperones in control of protein disaggregation. *EMBO J.* **27**, 328–335
57. Wise, M. J., and Tunnacliffe, A. (2004) POPP the question: What do LEA proteins do? *Trends Plant Sci.* **9**, 13–17
58. Zhang, F. X., Hu, Y., Huang, P., Toleman, C. A., Paterson, A. J., and Kudlow, J. E. (2007) Proteasome function is regulated by cyclic AMP-dependent protein kinase through phosphorylation of Rpt6. *J. Biol. Chem.* **282**, 22460–22471
59. Marshall, R. S., and Vierstra, R. D. (2019) Dynamic regulation of the 26S proteasome: From synthesis to degradation. *Front. Mol. Biosci.* **6**, 40
60. Bassham, D. C. (2007) Plant autophagy-more than a starvation response. *Curr. Opin. Plant Biol.* **10**, 587–593
61. Toyooka, K., Okamoto, T., and Minamikawa, T. (2001) Cotyledon cells of Vigna mungo seedlings use at least two distinct autophagic machineries for degradation of starch granules and cellular components. *J. Cell Biol.* **154**, 973–982
62. Vanderwilden, W., Herman, E. M., and Chrispeels, M. J. (1980) Protein bodies of mung bean cotyledons as autophagic organelles. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 428–432
63. Muntz, K. (1998) Deposition of storage proteins. *Plant Mol. Biol.* **38**, 77–99
64. Tosi, P., Parker, M., Gritsch, C. S., Carzaniga, R., Martin, B., and Shewry, P. R. (2009) Trafficking of storage proteins in developing grain of wheat. *J. Exp. Bot.* **60**, 979–991
65. Perez-Riverol, Y., Csordas, A., Bai, J. W., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Perez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., *et al*. (2019) The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450