Genetic characterization of cysteine-rich type-b avenin-like protein coding genes in common wheat

X. Y. Chen1,2,3,*, X. Y. Cao3,*, Y. J. Zhang2, S. Islam2, J. J. Zhang3, R. C. Yang2, J. J. Liu3, G. Y. Li3, R. Appels2, G. Keeble-Gagnere2, W. Q. Ji1, Z. H. He4 & W. J. Ma2

The wheat avenin-like proteins (ALP) are considered atypical gluten constituents and have shown positive effects on dough properties revealed using a transgenic approach. However, to date the genetic architecture of ALP genes is unclear, making it impossible to be utilized in wheat breeding. In the current study, three genes of type-b ALPs were identified and mapped to chromosomes 7AS, 4AL and 7DS. The coding gene sequence of both TaALP-7A and TaALP-7D was 855 bp long, encoding two identical homologous 284 amino acid long proteins. TaALP-4A was 858 bp long, encoding a 285 amino acid protein variant. Three alleles were identified for TaALP-7A and four for TaALP-4A. TaALP-7A alleles were of two types: type-1, which includes TaALP-7A1 and TaALP-7A2, encodes mature proteins, while type-2, represented by TaALP-7A3, contains a stop codon in the coding region and thus does not encode a mature protein. Dough quality testing of 102 wheat cultivars established a highly significant association of the type-1 TaALP-7A allele with better wheat processing quality. This allelic effects were confirmed among a range of commercial wheat cultivars. Our research makes the ALP be the first of such genetic variation source that can be readily utilized in wheat breeding.

Bread wheat (Triticum aestivum L.) is the most important staple food worldwide. Its unique viscoelastic properties conferred by the storage proteins, glutenins and gliadins, account for its extensive use and multi-ethnic expressions in food preparation, reflected in a wide range of wheat-derived food products1,2. The glutenin polymers are composed of an elastic backbone, formed by high-molecular-weight (HMW) subunits, and the branches, formed by low-molecular-weight (LMW) subunits which are the main contributors to dough strength and elasticity. The monomeric gliadins, conferring dough tractability, interact with the polymeric glutenins by strong covalent and non-covalent forces3–5. The structural characteristics of proteins affect polymerization behavior through both the strategic positioning of generally conserved cysteine residues and the presence of glutamine-rich repetitive regions within the polypeptide chain1–9. Cysteines constitute only a small proportion of the amino acids of gluten proteins (about2%), yet are extremely important to the structure and functionality of gluten due to their capacity to form intra- and inter-chain disulfide bonds10. Non-covalent bonds (hydrogen bridges, ionic interactions, and hydrophobic bonds), characteristically formed by gliadins, are responsible for the aggregation and structural stability of gluten proteins and dough structure formation11,12.

Besides the typical gluten proteins, storage protein components also include LMW gliadins or globulins with a molecular weight below 30,000 dalton13–15. Most of these atypical gluten proteins fall into the categories of ALPs or proteins with sequences similar to the previously reported LMW gliadin monomers14,15. LMW gliadins differ from gliadins and glutenins in lacking repetitive domains, with only a short sequence of proline and glutamine

1College of Agronomy, Northwest A & F University, Yangling 712100, Shaanxi, China. 2Australia-China Joint Centre for Wheat Improvement, School of Veterinary & Life Sciences, Murdoch University, Perth WA 6150, Australia. 3Cropp Research Institute, Shandong Academy of Agricultural Sciences/National Engineering Laboratory for Wheat and Maize/Key Laboratory of Wheat Biology and Genetic Improvement in North Yellow and Huai River Valley, Ministry of Agriculture, 250100, Jinan China. 4National Wheat Improvement Centre, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, 12 Zhongguancun South St, Haidian District, Beijing, China 100081. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to W.Q.J. (email: jiwanquan2003@126.com) or Z.H.H. (email: zhhecaas@163.com) or W.J.M. (email: w.ma@murdoch.edu.au)
residues present in the mature protein\textsuperscript{15}. The existence of proteins related to LMW gliadins, and constituting a new family of grain prolamin proteins, has also been confirmed in barley\textsuperscript{16,17} and rye\textsuperscript{18}. DuPont and co-workers described a protein isolated from wheat grains as ‘avenin-like’, based on partial amino sequences determined by mass spectrometry\textsuperscript{19}. Similarly, Vensel and coworkers\textsuperscript{20} identified five avenin-like proteins in the proteome of albumins and globulins during early and late stages of grain development. Kan and coworkers reported two classes of cDNAs encoding two types of ALPs, namely type-a and type-b\textsuperscript{21}. In a phylogenetic analysis of the prolamin superfamily, the ALP genes co-locate as a single cluster, with its closest neighbors being avenin of oats and the sulphur-rich proteins (\(\alpha\)-gliadins, \(\gamma\)-gliadins, LMW subunits of glutenin). Furthermore, in the same study the authors observed that type-a ALPs contain 14 cysteine residues, among which eight cysteines form the characteristic conserved cysteine skeleton of the typical prolamins (similar to avenin, \(\alpha\)-gliadins, \(\gamma\)-gliadins, and LMW subunits of glutenin)\textsuperscript{22}. It is noteworthy that type-a ALPs can form seven intra-chain disulfide bonds, which is typical of monomeric LMW gliadins. Type-b ALPs contain two repetitive domains (R1, R2), each with eight cysteine residues in homologous positions to the cysteines of \(\gamma\)-gliadin and oats avenin protein. Type-b ALPs also exhibited some differences in cysteine distribution, with a total of 18 or 19 cysteine residues. In particular, Mamone and coworkers\textsuperscript{23} detected type-b ALPs in the glutenin fraction of durum wheat cultivar Svevo, while Kan and co-workers\textsuperscript{21} found that the two cysteines in the N-terminal domain are not conserved in various Aegilops species, hence suggesting that they could be involved in inter-chain linkages to polymeric subunits of glutenins. The identification of type-b ALPs was supported by the acquisition of sequences from a reasonable number of tryptic peptides matching the expected molecular weights and pI values\textsuperscript{23}. The higher number of cysteines in type-b ALPs was expected to have a significant effect on folding and the arrangement of disulfide linkages, not only by stabilizing the molecular structure, but also by influencing glutenin polymer formation. Chen and coworkers\textsuperscript{24} predicted that type-b ALPs were capable of forming eight intra-molecular disulfide bonds, with three free cysteine residues involved in inter-molecular disulfide bond formation. They confirmed that type-b ALPs can notably perform as “chain branches”, increasing the probability of glutenin macro-polymer (GMP) formation and including other glutenin subunits\textsuperscript{24}. Ma and coworkers overexpressed type-b ALPs in two transgenic wheat lines, resulting in a highly significant improve of dough mixing properties and provided strong evidence for their incorporation into gluten polymers\textsuperscript{25}.

Until now, when selecting for dough and baking quality improvement, wheat breeders have mainly relied on the genetic variation underlying gluten proteins. The effect on dough mixing properties associated with ALPs represents a novel genetic effect that has not been utilized in a targeted way in wheat grain functionality breeding thus far. Marker-assisted selection targeting ALPs depends on both natural allelic variation of ALPs and their validated effects on dough mixing properties. The objectives of this study were to locate the type-b ALP coding genes, find the number of available alleles and quantify the allelic effects for each locus, and to develop allele-specific markers for wheat grain functionality breeding.

Results

Type-b ALP coding genes in \textit{Triticum aestivum}. ALP specific primers were used to amplify the complete coding sequence of type-b ALP genes from the genomic DNA of 19 cultivars. The amplified products covering the start and stop codons were about 902 bp in length (Fig. 1). Nucleotide sequences highly similar (99\%) to the previously reported type-b ALP gene sequence (Accession No. FJ529695) were obtained.

Sequence alignment and analysis. The sequences of the amplified ALP genes were used to search the EnsemblPlants (http://plants.ensembl.org/Triticum_aestivum/Info/Index) and the International Wheat Genome Sequencing Consortium (IWGSC) databases. The results showed that type-b ALP genes were transcribed at a
high rate and consisted of a single uninterrupted exon. The results were consistent with previous studies. In addition, the type-b ALP gene sequences were good matches to three surveyed sequences (Chinese Spring) on chromosomes 7DS (99%), 4AL (98%) and 7AS (97%).

**Gene locations.** Three pairs of specific primers were designed, targeting ALP genes on chromosomes 7AS, 4AL and 7DS to verify the blasted results (Table 1). These primers were tested across the entire set of Chinese Spring deletion lines. Results were consistent with the surveyed sequence databases and the chromosomal location of the gene products ALP-7A, ALP-4A and ALP-7D was confirmed (Fig. 2). We thus named the three ALP gene loci TaALPb-7A, TaALPb-4A, and TaALPb-7D, accordingly.

**SNP and indel analyses.** Genomic DNA of 19 cultivars was amplified using the primer pairs specific for ALP-7A, ALP-4A and ALP-7D, with each primer pair amplifying one single sequence across all cultivars. The full-length sequences at the three gene loci were either 855 or 858 bp, encoding proteins with 284 and 285 amino acid residues, accordingly. In addition, SNP and indel polymorphisms were discovered among the amplicons of different cultivars at loci TaALPb-7A and TaALPb-4A. Seven polymorphic sites were detected among the TaALPb-7A amplicons, including one deletion (three bases) and six SNPs involving five transversions and one transition (Fig. 3). Eighteen polymorphic sites were detected among the TaALPb-4A amplicons, including seventeen SNPs involving 14 transversions and 3 transitions, as well as one indel variant (Fig. 4). No variation was found at the TaALPb-7D locus (Fig. 5). These results indicate that multiple alleles exist for TaALPb-7A and 4A while no or little genetic variation exists at the TaALPb-7D locus. Further comparison revealed that the TaALPb-7A gene had three alleles, designated TaALPb-7A1 (GenBank accession no. KU286147), TaALPb-7A2 (GenBank accession no. KU286148) (frequency 50.98%) and TaALPb-7A3 (GenBank accession no. KU286149) (frequency 49.02%) in the current study (Table 2), while TaALPb-4A gene had four alleles, TaALPb-4A1 (GenBank accession no. KU286150), TaALPb-4A2 (GenBank accession no. KU286151), TaALPb-4A3 (GenBank accession no. KU286152), and TaALPb-4A4 (GenBank accession no. KU286153). The TaALPb-7D (GenBank accession no. KU286154) locus did not show any allelic variation across the cultivars screened in this study. Analysis of the translated protein sequences revealed that the signal peptides at the N- and C-termini were rather conserved, with hardly any variation detected. The sequence differences occurred mainly in the repetitive region. Major variations were detected on 7AS and 4AL alleles. Among the 7AS alleles, TaALPb-7A1 and TaALPb-7A2 encode mature proteins, while allele TaALPb-7A3 contained a stop codon (a SNP resulting in CAA → TAA codon change. Figure 3), leading to early termination of translation in 10 cultivars (Supplementary Figure 1). Anonymous silenced ALP genes have been previously reported. In-frame stop codons were not detected for the 4AL alleles, although many variations occurred within the mature proteins (Supplementary Figure 2). In addition, 18 cysteine residues

| Marker | Primer sequences (5′-3′) | Tm (°C) | Product size (bp) | Amplified target |
|--------|--------------------------|---------|------------------|-----------------|
| ALP    | F: TGCCACACATGATGATGCATG | 60      | 912              | Full length     |
|        | R: ATGAAGGTCTTCATCCTGGCTC |         |                  |                 |
| ALP-7A | F: ATGCCAATCAACACAGG     | 55      | 762              | 7A genome-specific |
|        | R: TGTAGCCACACCAGGTTAA   |         |                  |                 |
| ALP-4A | F: TGGGACATCAACACAACAG   | 55      | 777              | 4A genome-specific |
|        | R: TCTAGCATGCACACTAGTGC  |         |                  |                 |
| ALP-7D | F: ATGAAGGTCTTCATCCTGGCT | 58      | 805              | 7D genome-specific |
|        | R: CTGCAGCGCAACACAGT     |         |                  |                 |

Table 1. Chromosome-specific primer sets for cloning type-b ALP genes.

![Figure 2. Chromosome-specific type-b ALP amplification using Chinese Spring deletion lines.](image-url)
were detected in 7AS and 7DS ALPs. The 4AL ALPs contained 19 cysteine residues, exhibiting more cysteine residues than previously reported for endosperm-specific storage proteins.

In general, the type-b ALP proteins can be considered to be glutamine and proline-rich proteins, although less than gliadins and LMW glutenins, due to the lack of extensive repetitive sequences. At the same time, ALP proteins exhibited a conserved distribution of cysteines (Supplementary Figures 1 and 2), which are predicted to be able to form seven or eight intra-molecular disulfide bonds among the 18 or 19 cysteine residues. The remaining cysteines (at least two) may form inter-molecular disulfide bonds linking to adjacent storage protein subunits.

Phylogenetic analysis. The phylogenetic relationship of the 42 cloned type b ALP sequences was analyzed by applying UPGMA to the aligned complete coding sequences of all clones and wheat storage protein genes, as well as the reported ALPs of wheat-related species available from various databases (Table 3). As shown in Fig. 6, the cloned type-b ALP sequences clustered according to their chromosomal origin. The cloned ALP sequences were closest to the reported type-b ALP sequences of related species, followed by sequences corresponding to HMW-GS and LMW-GS, while ω-gliadin were the furthest in evolutionary terms (Fig. 6).
Allelic effects. The fact that the TaALPb-7A locus has two types of alleles, active and silent, allowed us to study its allelic effects. Allele-specific PCR markers were designed to differentiate the two types of TaALPb-7A alleles and a total 102 wheat cultivars or lines were selected for quality testing and marker analysis (Fig. 7).
Mixograph analyses were conducted to assess wheat dough strength using procedures published previously. Significant allelic effect differences were detected between the active and silent alleles of TaALPb-7A. The active allele was significantly associated with higher dough strength parameters, including Midline Peak Time (P < 0.0443), and Midline Time × Width (P < 0.0096) (Table 2). Meanwhile, the component of HMW-GS, protein content and gluten content of 102 wheat cultivars or lines were analyzed (see Supplementary Table 1). Results revealed that the HMW-GS alleles were randomly distributed between the allelic types. The favorable subunit Dx5 was found in 33% of the silent wheat lines and 29% in the active lines. No significant association was detected between the allelic types and grain protein content or gluten content (Table 4), indicating the high dough strength of the active allelic type is from the expression of TaALPb-7A.

**Figure 6.** Phylogenetic analysis of the cloned sequences of type-b ALPs.

**Figure 7.** PCR amplification products of AS-PCR marker for TaALP-7A. M DNA ladder 100bp; 01 to 11 lanes are PCR products of selected cultivars; 12, Negative control; presence of a band indicates TaALP-7A1 and TaALP-7A2, absence stands for TaALP-7A3.
Comparison of the active and silent alleles of TaALPb-7A at gene expression level. To further confirm the function of these two types of the TaALPb-7A alleles, comparison of gene expression between four Australian cultivars (Kauz, Yitpi, Gregory, and Chara) containing the active TaALPb-7A allele and another four wheat cultivars (Chinese Spring, Eagle Rock, Westonia, and Wyalketchem) containing the silent TaALPb-7A allele were conducted by using reverse transcription (RT) reaction followed by digital droplet PCR (ddPCR). Results revealed that the cultivars with the active allele give a normal gene expression, the ratios of expressed gene copy numbers between TaALPb-7A and actin ranged from 1:2.54 to 1:3.36 (Table 5), while the four cultivars with the silent allele had no gene expression.

Discussion
Cysteine-rich wheat grain storage avenin-like proteins (ALPs) capable of forming intra-molecular disulfide bonds were discovered in recent years and are considered atypical gluten components of the wheat grain storage protein complement. However, the presence of similar low-molecular weight subunits in glutenins and gliadins has been reported in the 1970s\(^35,36\) and these seem capable of forming strong in vivo associations among themselves and with HMW-GS and LMW-GS, apparently by inter-chain disulfide bonds. ALPs make up about 1% of total endosperm proteins\(^37,38\). Contrary to the typical gluten proteins that are characterized by large repetitive central domains these non-traditional gluten proteins lack repeating sequences. In 2D gels, type-b ALPs migrate only slightly faster than the LMW glutenins, α-, or γ-gliadins, due to sequence duplication in the central domain (R1, R2), compensating to a large extent for the missing repeating sequence domain\(^37,38\). The unique properties demonstrated by type-b ALPs make them an ideal component of elastic disulfide-linked aggregates. In this study, phylogenetic analysis clearly showed that the type-b ALP sequences of common wheat clustered in the same
Aegilops spp. In contrast, the sequences of type-b ALP genes indicated a genetic relationship to the unique C-terminal domains of gluten proteins (LMW-GS and gliadins) and are notable for the absence of significant repetitive domains of typical HMW-GS. The tetraploid genome of Aegilops and hexaploid wheat of Triticum species provides a theoretical basis for the local variation at the 7D locus of type-b ALP genes in the lines and varieties investigated in this study.

Despite the potential of type-b ALP proteins to form intermolecular bonds, their low abundance and the absence of a repetitive domain might limit their ability to play a major role in determining dough functional properties. Research conducted on transgenic type-b ALP wheat lines confirmed the presence of free cysteines capable of forming extra inter-chain disulfide linkages with glutenins (HMW-GS) in the presence of sulfur-rich proteins (LMW-GS and gliadins). Combination and association analysis using targeted allelic ALP combinations will shed further light upon the highly complex interactions due to the allelic composition of sulfur-rich proteins (γ- and α-gliadin, LMW-GS), as well as sulfur-poor proteins (ω-gliadins and HMW-GS).

Although many researchers have mentioned that type-b ALP genes of wheat belong to a multigene family, there still remained a paucity of genetic information about the chromosomal location, number of loci and alleles, and allelic effects. The current study clearly identified the chromosomal locations of type-b ALP genes and the number of alleles at each locus for the first time. In this study, the three type-b ALP gene loci were mapped to chromosomes 7AS, 4AL and 7DS. Theoretically, due to the allohexaploid (AABBDD) nature of bread wheat, the three gene loci should be located on three homeologous chromosome locations (7AS, 7BS and 7DS). The reason for the unusual chromosomal locations can be found in the evolutionary relationships of wheat chromosomes arms, i.e., a 4AL/7BS translocation, a pericentric inversion, and a paracentric inversion that took place in the tetraploid progenitor of hexaploid wheat. This clearly provides a theoretical basis for identifying the potential of individual ALPs for dough viscoelasticity improvement. Research conducted on transgenic type-b ALP wheat lines confirmed the presence of free cysteines capable of forming inter-chain disulfide linkages with glutenins (HMW-GS and LMW-GS). Future research on the expression of ALPs, aimed at developing a more detailed understanding of peptide chain interactions, disulfide bond arrangements, and tertiary structure formation will allow us to delve deeper into the molecular interactions with gluten proteins. Combination and association analysis using targeted allelic ALP combinations will shed further light upon the highly complex interactions due to the allelic composition of sulfur-rich proteins (γ- and α-gliadin, LMW-GS), as well as sulfur-poor proteins (ω-gliadins and HMW-GS).

Methods

Plant material and experimental design. All wheat lines used in the current study is listed in Table 6. Nineteen wheat cultivars from Australia and China were used to clone the type-b ALP genes. Field trial of 102 bread wheat cultivars (lines) with a randomized complete block design with 3 replications at the experiment station in Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China, in 2012 and 2013 (36°42′N, 117°4′E; altitude 48 m). Shandong has a humid subtropical climate with a mean annual rainfall of 700 mm and average maximum temperatures of over 34 °C during wheat growing season. Seeds were sown by 300 kernels per square meter in 4 × 6 m plots.
Fertilizers of 120 kg N ha$^{-1}$, 60 kg P$_2$O$_5$ ha$^{-1}$, and 120 kg K$_2$O ha$^{-1}$ were applied to the soil prior to sowing, and another 120 kg N ha$^{-1}$ was top dressed at jointing in accordance with local wheat farming practices. The soil contained 1.64 g kg$^{-1}$ organic matter in both years. All other standard agronomical practices were adopted. Seeds were harvested and used for mixograph, HMW-GS and NIR analysis. The means values of two trials were taken for analysis.

### DNA extraction and PCR amplification.
Genomic DNA of 19 cultivars was extracted from 1-week-old seedlings by using the cetyltrimethyl ammonium bromide (CTAB) method$^{51}$. Primers to amplify the full-length gene were designed based on the type-b ALP coding sequence from NCBI database (Accession No.FJ529695) (Table 2). The PCR conditions were set to 95 °C for 5 min, 35 cycles of 95 °C for 30 s, 60 °C for 45 s and 72 °C for 50 s, and a final extension at 72 °C for 10 min. PCR products were separated by 1.5% (w/v) agarose gel electrophoresis, and the expected fragments were purified from the gel using a Gel Extraction Kit (Promega, Madison, WI, USA). Subsequently, the purified PCR products were amplified using BigDye@version 3.1 terminator mix (Applied Biosystems) and submitted for Sanger sequencing at the Western Australia State Agricultural Biotechnology Centre. PCR and DNA sequencing were repeated three times to ensure the accuracy.

### Chromosomal locations of type-b ALP genes.
The EnsemblPlants (http://plants.ensembl.org/Triticum_aestivum/Info/Index) and the International Wheat Genome Sequencing Consortium (IWGSC) databases were used to analyze the obtained sequences. After blasting the obtained sequence from 19 cultivars against the databases, good matches were found on chromosomes 7DS, 4AL and 7AS. Based on this, three pairs of specific primers were designed for each chromosome using Primer V5.0 software (http://www.premierbiosoft.com) (Table 2). Chinese Spring deletion lines were then used to test the designed chromosome-specific primers and to verify the chromosomal locations.

### Sequence analysis.
The chrome-specific primers were used to amplify the genomic DNA of 19 cultivars. The PCR products were ligated into pGEM-T Easy vector (Promega, Madison, WI, USA) following the manufacturer’s protocol and then the hybrid vector was transformed into competent cells of E. coli strain DH-5$\alpha$. Plasmids were extracted using the Magic Mini Plasmid Prep kit (Promega, Madison, WI, USA) and the extracted DNA was amplified using BigDye® version 3.1 terminator mix (Applied Biosystems) for Sanger sequencing. The program Bioedit 7.0 was used for sequence analysis. Geneious® software (R7) was used for multiple alignment of the translated amino acid sequences and phylogenetic analysis.

### Table 6. Name and origin of 111 wheat cultivars and advanced lines.

| Name      | Origin     | Name      | Origin     | Name      | Origin     | Name      | Origin     |
|-----------|------------|-----------|------------|-----------|------------|-----------|------------|
| JimaiH101 | CHINA      | JimaiT112 | CHINA      | JimaiC70218 | CHINA     | Jimai13J390 | CHINA     |
| JimaiH102 | CHINA      | JimaiT118 | CHINA      | JimaiC70223 | CHINA     | Jimai13P406 | CHINA     |
| JimaiH105 | CHINA      | JimaiT120 | CHINA      | JimaiC70228 | CHINA     | Jimai13J407 | CHINA     |
| JimaiH106 | CHINA      | JimaiT101 | CHINA      | JimaiC70231 | CHINA     | Jimai13J408 | CHINA     |
| JimaiH107 | CHINA      | JimaiT102 | CHINA      | JimaiC70241 | CHINA     | Jimai13J492 | CHINA     |
| JimaiH108 | CHINA      | JimaiT103 | CHINA      | JimaiC70245 | CHINA     | Jimai13J424 | CHINA     |
| JimaiH109 | CHINA      | JimaiT104 | CHINA      | JimaiC70247 | CHINA     | Jimai13J427 | CHINA     |
| JimaiH110 | CHINA      | JimaiT105 | CHINA      | JimaiC70285 | CHINA     | Jimai13J394 | CHINA     |
| JimaiH111 | CHINA      | JimaiT106 | CHINA      | JimaiC70298 | CHINA     | Jimai13J464 | CHINA     |
| JimaiH112 | CHINA      | JimaiT107 | CHINA      | JimaiC70321 | CHINA     | Jimai13J467 | CHINA     |
| JimaiH113 | CHINA      | JimaiT108 | CHINA      | JimaiC70356 | CHINA     | Jimai13J490 | CHINA     |
| JimaiH114 | CHINA      | JimaiT109 | CHINA      | JimaiC70361 | CHINA     | Jimai13J492 | CHINA     |
| JimaiH117 | CHINA      | JimaiT111 | CHINA      | JimaiC70365 | CHINA     | Jimai13J494 | CHINA     |
| JimaiH118 | CHINA      | JimaiT113 | CHINA      | JimaiC70373 | CHINA     | Jimai13J495 | CHINA     |
| JimaiH120 | CHINA      | JimaiT116 | CHINA      | JimaiC70421 | CHINA     | Jimai9088  | CHINA     |
| JimaiH122 | CHINA      | JimaiT117 | CHINA      | JimaiC70445 | CHINA     | Jimai23    | CHINA     |
| JimaiH123 | CHINA      | JimaiT118 | CHINA      | JimaiC70459 | CHINA     | Jimai24    | CHINA     |
| JimaiH124 | CHINA      | JimaiT119 | CHINA      | JimaiC70483 | CHINA     | Jimai10860229 | CHINA   |
| JimaiH125 | CHINA      | JimaiT120 | CHINA      | JimaiC70509 | CHINA     | Chinese Spring | CHINA     |
| JimaiT102 | CHINA      | JimaiT121 | CHINA      | JimaiT30005 | CHINA     | Kauz      | AUSTRALIA |
| JimaiT103 | CHINA      | JimaiT122 | CHINA      | JimaiT40097 | CHINA     | Eagle Rock | AUSTRALIA |
| JimaiT104 | CHINA      | JimaiT123 | CHINA      | JimaiT40098 | CHINA     | Chara      | AUSTRALIA |
| JimaiT105 | CHINA      | JimaiT124 | CHINA      | JimaiT40103 | CHINA     | Wyalkatchem | AUSTRALIA |
| JimaiT216 | CHINA      | Jimai13P307 | CHINA    | JimaiT40271 | CHINA     | Gregory    | AUSTRALIA |
| JimaiT108 | CHINA      | Jimai13P406 | CHINA    | JimaiT40284 | CHINA     | Living Stone | AUSTRALIA |
| JimaiT109 | CHINA      | Jimai44   | CHINA      | JimaiT40362 | CHINA     | Yitpi      | AUSTRALIA |
| JimaiT110 | CHINA      | Jimai13P414 | CHINA    | JimaiT40368 | CHINA     | Westonia   | AUSTRALIA |
| JimaiT111 | CHINA      | JimaiC70107 | CHINA    | Jimai13J386 | CHINA     | —         | —         |
Allele-specific marker development. Primers targeting the type-1 TaALP-7A allele were designed based on SNP/InDel information: F: 5’-TGCAAGCAGCTCTAGCCCTCAG-3’; R: 5’-GTGCTGTT AGGCTGATCCACCAGGA-3’. A total of 102 wheat cultivars and lines (Table 1) were screened using the allele-specific primers.

HMW-GS electrophoretic analysis. The HMW-GS protein for SDS-PAGE was extracted from wheat grains by using a modified method based on Singh et al. In detail, 500μl of 55% (v/v) isopropanol was mixed with crushed individual kernels for 5 min through continuous vortexing, followed by incubation (30 min at 65°C), vortexing (5 min), and centrifugation (5 min at 10000 rpm). This step was repeated three times to completely remove gliadins. Add 600μl of 62.5 mM Tris-HCl (pH 6.8) buffer containing 10% (v/v) glycerol, 2% (w/v) sodium dodecyl sulfate (SDS), 0.003% (w/v) bromophenol blue, and 5% β-mercaptoethanol. The samples were boiled for 2 hours and then centrifuged for 5 minutes at 10000 rpm, 15 ml of upper solution of each sample were loaded on to the gel. Proteins were separated by SDS-PAGE according to Jackson et al.35 using stacking separation gels containing 4% acrylamide, 0.3% bis acrylamide, 0.1% SDS, and 0.125 M Tris-HCL (pH 6.8), and 8.7% acrylamide, 0.3% bis acrylamide, 0.1% SDS, and 0.38 M Tris-HCL (pH 8.8). The bands of HMW-GS on SDS-PAGE were scored according to the nomenclature system described by Payne and Lawrence34.

Quality testing. A 10-gram mixograph (National Manufacturing Co., Lincoln NE) was used to evaluate wheat dough mixing properties, as described by Zhang and coworkers35. Mixograph Peak Time (MPT, min), Peak Integral (MPI, cm²), Peak Width (MPW, %), and Midline Time × Width (MT × W, min) were measured as the four parameters selected for evaluating the dough quality. The statistical significance of mixograph data was assessed performing T-tests using the SAS/STAT System software, Version 8.0 (SAS Institute Inc. Cary, NC)35. DA7200 near infrared apparatus (Perten, Swedish) was applied to analyze the protein content and gluten content following the manufacture’s suggestion.

RNA extraction and ddPCR. RNA was extracted from 2 mature wheat grains of the 8 cultivars: Kauz, Yitpi, Gregory, Chara, Eagle spring, Asia, and Wyalketchem using the Qiagen RNaseasy mini kit. The TaALP-7A RNA was carried out using the SuperScript® II Reverse Transcriptase (Applied Biosystems), with the 3′ primer 5′-GCTGGTAGGCTGATCCACCAGGA-3′ for the active allele and 3′ primer 5′-GCTGGTG AGGCTGATCCACCAGT-3′ for the silent allele. The ddPCR was performed in a QX200 ddPCR system (Bio-Rad). The forward primers were 5′-TGCAGCAGCTTAGCTGCTGCCAT-3′ for the active allele and 5′-TGGACGAGCTTAGCTGCTGCCAG-3′ for the silent allele. The beta actin primers (F: 5′- AGACCTAGAGCTGCTGCCAG-3′; R: 5′-ACACTTGTTGGCTACAG-3′) were used as the reference gene in a separate ddPCR. The ratios of the expressed gene copies between actin and TaALP-7A were calculated.

References
1. Shewry, P. R., Napier, J. A. & Tatham, A. S. Seed Storage Proteins: Structures and Biosynthesis. Plant Cell. 7, 945–956 (1995).
2. Shewry, P. R. & Tatham, A. S. Biotechnology of Wheat Quality. J. Cereal Sci. 73, 397–406 (1997).
3. Shewry, P. R. Wheat. J. Exp. Bot. 60, 1537–1553 (2009).
4. Payne, P. I. Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annu. Rev. Plant Biol. 38, 141–153 (1987).
5. Shewry, P. R. & Tatham, A. S. Disulphide bonds in wheat gluten proteins. J. Cereal Sci. 25, 207–227 (1997).
6. Kohler, P., Bertl, H.-D. & Wieser, H. Disulphide bonds in wheat gluten: further cystine peptides from high molecular weight (HMW) and low molecular weight (LMW) subunits of glutenin and from γ-gliadins. Z Lebensm Unters Forsch. 196, 237–249 (1993).
7. Shewry, P. R. & Tatham, A. S. Disulphide Bonds in Wheat Gluten Proteins. J. Cereal Sci. 25, 207–227 (1997).
8. Shewry, P. R. & Tatham, A. S. Disulphide bond in wheat gluten proteins. J. Cereal Sci. 25, 207–227 (1997).
9. Shewry, P. R. & Halford, N. G. Cereal seed storage proteins: structures, properties and role in grain utilization. J. Exp. Bot. 53, 947–958 (2002).
10. Grosch, W. & Wieser, H. Redox Reactions in Wheat Dough as Affected by Ascorbic Acid. J. Cereal Sci. 29, 1–16 (1999).
11. Wriley, C., Békés, F. & Bushuk, W. Gliadin and glutenin: the unique balance of wheat quality. (American Association of Cereal Chemists, Inc (AACCC) 2006).
12. Schiraldi, A., Piazza, L., Fessas, D. & Riva, M. I. Handbook of thermal analysis and calorimetry from macromolecules to man. Vol. 4, 829–921 (Amsterdam: Elsevier, 1999).
13. Salcedo, G., Prada, J. & Aragoncillo, C. Low MW gluten-like proteins from wheat endosperm. Phytochem. 18, 725–727 (1979).
14. Anderson, O. D., Hsiia, C. C., Adalsteins, A. E., Lew, E. L. & Kasarda, D. D. Identification of several new classes of low-molecular-weight wheat gliadin-related proteins and genes. Theor. Appl. Genet. 103, 307–315 (2001).
15. Clarke, B. C., Aragoncillo, C. & Salcedo, G. Two groups of low-molecular weight hydrophobic proteins from barley endosperm. J. Exp. Bot. 32, 1279–1286 (1981).
16. Salcedo, G., Sanchez-Monge, R., Argamenteria, A. & Aragoncillo, C. Low-molecular-weight prolamins–purification of a component from barley endosperm. J. Agric. Food Chem. 30, 1155–1157 (1982).
17. Schiraldi, A., Calero, M., Soriano, F. & Mirnez, E. Identification of major rye secalins as coeliac immunoreactive proteins. Biochim Biophys Acta - Protein Structure and Molecular Enzymology. 1295, 13–22 (1996).
18. DuPont, F. M., Chan, R., Lopez, R. & Vensel, W. H. Sequential extraction and quantitative recovery of gliadins, glutenins, and other proteins from small samples of wheat flour. J. Agric. Food Chem. 53, 1575–1584 (2005).
19. Shewry, P. R., Jenkins, J., Beaudoin, F. & Mills, E. C. The classification, functions and evolutionary relationships of plant proteins in relation to food allergens. Plant food allergens, 24–41 (2004).
20. Marmone, G., Caro, S. D., Luccia, A. D., Addo, F. & Ferranti, P. Proteomic–based analytical approach for the characterization of glutenin subunits in durum wheat. J. Mass Spectrom. 44, 1789–1793 (2009).
24. Chen, P., Li, R., Zhou, R., He, G. & Shewry, P. R. Heterologous expression and dough mixing studies of a novel cysteine-rich Avenin-like protein. Cereal Res Commun. 38, 406–418 (2010).
25. Ma, F. Y. et al. Overexpression of avenin-like b proteins in bread wheat (Triticum aestivum L.) improves dough mixing properties by their incorporation into glutenin polymers. PLoS one. 7, 1–11 (2013).
26. Chen, P. et al. Cloning, expression and characterization of novel avenin-like genes in wheat and related species. J. Cereal Sci. 48, 734–740 (2008).
27. Anderson, O. D. & Greene, F. C. The o-gladien gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor. Appl. Genet. 95, 59–65 (1997).
28. Beier, H. H. & Grimm, M. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res. 29, 4767–4782 (2001).
29. Santos, M. A., Moura, G., Massey, S. E. & Tuite, M. F. Driving change: the evolution of alternative genetic codes. Trends in Genetics. 20, 95–102 (2004).
30. Lorenzo, A. & Kronstad, W. F. Reliability of Two Laboratory Techniques to Predict Bread Wheat Protein Quality in Nontraditional Growing Areas. Crop Sci. 27, 247–252 (1987).
31. Martinez, J. P. et al. Relationships Between Mixograph Parameters and Indices of Wheat Grain Quality. J. Cereal Sci. 27, 179–189 (1998).
32. Khatkar, B. S., Bell, A. E. & Schofield, J. D. A Comparative Study of the Inter-Relationships-Between Mixograph Parameters and Bread. J. Sci. Food Agr. 72, 71–85 (1996).
33. Bordes, J., Branlard, G., Oury, F. X., Charmet, G. & Balfourier, F. Agronomic characteristics, grain quality and flour rheology of 372 bread wheats in a worldwide core collection. J. Cereal Sci. 48, 569–579 (2008).
34. Ohm, J. B., Simsek, S. & Mergoum, M. Modeling of Dough Mixing Profile Under Thermal and Nonthermal Constraint for Evaluation of Breadmaking Quality of Hard Spring Wheat Flour. Cereal Chem. 89, 135–141 (2012).
35. Bietz, J. A. & Wall, J. S. Wheat gluten subunits: Molecular weights determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Cereal Chem. 49, 416–430 (1972).
36. Payne, P. I. & Corfield, K. G. Subunit composition of wheat gluten proteins, isolated by gel filtration in a dissociating medium. Planta. 145, 83–88 (1979).
37. Dupont, F. M., Vensed, W. H., Tanaka, C. K., Hurkman, W. J. & Altenbach, S. B. Deciphering the complexities of the wheat flour proteome using quantitative two-dimensional electrophoresis, three proteases and tandem mass spectrometry. Proteome Sci. 9–1186 (2011).
38. Kasarda, D. D. et al. Farinini: Characterization of a novel wheat endosperm protein belonging to the prolamin superfamily. J. Agr. Food Chem. 61, 2407–2417 (2013).
39. Cassidy, B. G., Dvorak, J. & Anderson, O. D. The wheat low-molecular weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. Theor. Appl. Genet. 96, 743–750 (1998).
40. Sabelli, P. A. & Shewry, P. R. Characterization and organization of gene families at the Gli-1 loci of bread and durum wheats by restriction fragment analysis. Theor. Appl. Genet. 83, 209–216 (1991).
41. Thompson, R. D., Bartels, D., Harberd, N. P. & Flavell, R. B. Characterization of the multigene family coding for HMW glutenin subunits in wheat using cDNA clones. Theor. Appl. Genet. 87, 87–96 (1983).
42. Cole, E., Fullington, J. & Kasarda, D. D. Grain protein variability among species of Triticum and Aegilops: quantitative SDS-PAGE studies. Theor. Appl. Genet. 60, 17–30 (1981).
43. Ma, F. Y. et al. Transformation of common wheat (Triticum aestivum L.) with avenin-like b gene improves flour mixing properties. Mol. Breed. 32, 853–865 (2013).
44. Dvorak, J. & Chen, K. C. Distribution of Nonstructural Variation between Wheat Cultivars along Chromosome Arm 6Bp: Evidence from the Linkage Map and Physical Map of the Arm. Genetics. 106, 325–333 (1984).
45. Dvořák, J., Terlizzi, P. D., Zhang, H. B. & Resta, P. The evolution of polyploid wheats: identification of the A genome donor species. Genome. 36, 21–31 (1993).
46. Naranjo, T., Roca, A., Goicoechea, P. G. & Giraldez, R. Arm homoeology of wheat and rye chromosomes. Genome. 29, 873–882 (1987).
47. Naranjo, T. Chromosome structure of durum wheat. Theor. Appl. Genet. 79, 397–400 (1990).
48. Devos, K. M., Dubcovsky, J., Dvořák, J., Chinoy, C. N. & Gale, M. D. Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor. Appl. Genet. 91, 282–288 (1995).
49. Andersen, J. R. Functional markers in plants. Trends in plant science. 8, 554–560 (2003).
50. Liu, Y. A., He, Z. H., Appels, R. & Xia, X. C. Functional markers in wheat: current status and future prospects. Theor. Appl. Genet. 125, 1–10 (2012).
51. Stacey, J. & Isaac, P. G. Isolation of DNA from plants. In Protocols for Nucleic Acid Analysis by Nonradioactive Probes. 9–15 (Humana Press, 1994).
52. Singh, N. K., Shepherd, K. W. & Cornish, G. B. A simplified SDS-PAGE procedure for separating LMW subunits of glutenin. J. Cereal Sci. 14, 203–208 (1991).
53. Jackson, E. A. et al. Proposal for combining the classification systems of alleles of Gli-1 and Glu-3 loci in bread wheat (Triticum aestivum L.). J Genet Breed. 50, 321–336 (1996).
54. Payne, P. I. & Lawrence, G. J. Catalogue of alleles for the complex 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 (1983).
55. Zhang, Y. et al. The gluten protein and interactions between components determine mixograph properties in an F6 recombinant inbred line population in bread wheat. J. Cereal Sci. 50, 219–226 (2009).

Acknowledgements
This work was financially supported by Australian Grain Research & Development Corporation project UMU00043 and Seed Industry Project of Taishan Scholar, Yong Talents Training Program of Shandong Academy of Agricultural Sciences.

Author Contributions
X.Y.C. performed the sequencing and RT-PCR experiments as well as wheat quality assessment and manuscript writing; W.J.M. conceived and designed the research; Y.Y.C. conducted the data analysis and part of the gene sequencing work; Y.J.Z. performed part of the sequencing and RT-PCR experiments; J.J.Z. helped sequencing and RT-PCR experiments as well as provide the chromosomal assigning material; R.C.Y. performed part of the RT-PCR experiments; J.J.L. carried out field trials and quality assessment; G.Y.L. conducted the wheat quality assessment; H.Z.H. conducted the data analysis and part of field trials; W.Q.J. conducted part of the data analysis and field trials; R.A. performed the genomic data analysis; S.I. supervised the daily operation of the research; G.K.-G. conducted the genomic data analysis.
Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Chen, X. Y. et al. Genetic characterization of cysteine-rich type-b avenin-like protein coding genes in common wheat. Sci. Rep. 6, 30692; doi: 10.1038/srep30692 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016