Genome-wide identification and evolutionary and expression analysis of the wheat amino acid transporter, an important gene family involved in abiotic stresses

Ruizheng Tian*,†, Yang Yangb,† and Maohua Chen*a,*

a Northwest A&F University, State Key Laboratory of Crop Stress Biology for Arid Areas, Key Laboratory of Integrated Pest Management on Crops in Northwestern Loess Plateau, Ministry of Agriculture and Rural Affairs, Yangling 712100, China.

b Northwest A&F University, State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Yangling 712100, China

* Correspondence: Northwest A&F University, State Key Laboratory of Crop Stress Biology for Arid Areas, Key Laboratory of Integrated Pest Management on Crops in Northwestern Loess Plateau, Ministry of Agriculture and Rural Affairs, Yangling 712100, China.

E-mail address: maohua.chen@nwsuaf.edu.cn (M. Chen); Tel: 862987082594

† These two authors contributed equally to this work.
Abstract

Background: Amino acid transporters (AATs), which transport amino acids across cellular membranes, play important roles in alleviating plant damage under stresses as well as in plant growth and development. Although this family has been systematically studied in many plant species, little is known about the AAT genes in bread wheat (*Triticum aestivum* L.) due to its complex genome sequence.

Results: In this study, a total of 296 AAT genes were identified from the latest wheat genome sequence (IWGSC v1.1) and classified into twelve distinct subfamilies based upon their sequence composition and phylogenetic relationship. Wheat AAT family members showed significant heterogeneity in chromosome distribution, with relatively high density in specific chromosomal regions. Comparison the number variation of gene copies and transmembrane regions of AAT genes in different sub-genome showed that the functional adaptation of the wheat AAT family during wheat polyploidization was driven mainly by sequence mutations rather than copy number variation. In addition, it was confirmed that changes in gene structure and protein conserved domains played important roles in the functional differentiation of the AAT family. Finally, the expression profiles of these TaAAT genes under heat, drought and salt stress and in the development stage of wheat showed that the expression of TaAATs exhibited abundant and distinct expression patterns under different abiotic stresses or in different tissues, and several important candidate AAT genes that may affect abiotic stress response and grain quality were also identified.

Conclusions: In this study, a total of 297 AAT proteins were systematically identified and characterized. Our study highlighted the important roles of gene duplication events in the expansion and functional differentiation of the wheat AAT family. The expression profiles of TaAATs revealed their importance for the grain development of wheat and their response to biotic and abiotic stresses. Our study also provided a theoretical basis for the further functional identification and utilization of the AAT gene family in wheat or other crops.

Keywords: wheat (*Triticum aestivum* L.); amino acid transporter; genome-wide; selective pressure; expression pattern; abiotic stress
**Background**

Amino acids are not only indispensable for the formation of proteins but also the main carrier for nitrogen exchange in plants [1, 2]. The obtainment of organic nitrogen depends to a large extent on the transport of amino acids, especially in processes of seed germination and seedling growth, mainly by amino acid transporters (AATs) [3, 4]. Moreover, studies have confirmed that AATs play important roles in determining the nutritional quality and protein content of seeds [5-7]. AATs are essential membrane proteins responsible for the transport of amino acids across cellular membranes in higher plants and their functions in multiple physiological processes of plant growth and development, including long-distance transport of amino acids, absorption of amino acids from soil and responses to pathogens and abiotic stresses [8-13]. The AAT gene family has been systematically identified and characterized in several plant species; for instance, there are 63 in *Arabidopsis*, 85 in rice, 189 in soybean, and 72 in potato, indicating that the AAT gene family is widely present in higher plants [14-17]. This gene family in plants includes the amino acid-polyamine-choline (APC) family and the amino acid/auxin permease (AAAP) family, both of which belong to the APC transporter superfamily. The APC family is mainly composed of three subfamilies - amino acid/choline transporters (ACTs), cationic amino acid transporters (CATs) and polyamine H+-symporters (PHSs), with tyrosine-specific transporters (TTPs) also identified in some plants [18, 19, 20]. The AAAP family contains at least six subfamilies, including lysine and histidine transporters (LHTs), γ-aminobutyric acid transporters (GATs), proline transporters (ProTs), amino acid permeases (AAPs), auxin transporters (AUXs), and aromatic and neutral amino acid transporters (ANTs) [18, 20, 21].

Although the AAT family has been identified by genome-wide scanning in multiple plant species, the most detailed characterization of its functions is mainly concentrated in *Arabidopsis*. Many AAT genes alleviate plant damage under water stress by promoting the transport of stress-related compounds and compatible solutes [22]. For example, *AtProTs* are responsible for the transport of various substrates, such as proline, glycine betaine and γ-aminobutyric acid (GABA), among which *AtProT1* can rapidly transport proline under water stress to reduce plant damage [23]. Although ProTs have similar subcellular localization and substrate specificity, in tomato, LeProT1 transports glycine and GABA with different affinities, while in *Arabidopsis*, *AtProTs* are specifically expressed in tissues with elevated proline content [23, 24]. Moreover, *AtProT2* also plays a crucial role in the transport of compatible solutes to the root tip region, as evidenced by studies in barley [23, 25]. In addition, the AAT genes also affect plant growth and development through the regulation of auxin [26]. For example, *AtAUX1* promotes the tropism of the roots and the formation of lateral roots by mediating the influx of auxin into the roots [26, 27].

In addition to alleviate plant damage under water stress by promoting the transport of stress-related compounds and compatible solutes, many AAT genes directly affect embryonic development and seed protein content. The *Arabidopsis* AAP subfamily contains eight members: *AtAAP1*-AtAAP8. Notably, detailed characterization of AtAAPs by heterologous expression systems has demonstrated that six AtAAPs preferentially transport neutral and charged amino acids with different specificities and affinities [28, 29]. *AtAAP1* is specifically expressed in cotyledons and endosperm, regulating the transport of amino acids to root cells or developing embryos, and is therefore essential for seed yield and storage protein synthesis [10, 30]. *AtAAP2* affects the transport of amino acids from xylem to phloem [31]. Analysis of the aap6 mutant confirmed that *AtAAP6* affects the interaction with aphids by
modulating the amino acid content in the sieve elements of Arabidopsis [20]. AtAAP8 is considered to play an important role in the uptake of amino acids in embryos and endosperms during early embryonic development [32]. In addition, the functions of AAP genes have been studied in other species and have been confirmed to be related to grain development and protein content [5-7]. For example, the content of storage proteins in seeds of Vicia narbonensis and Pisum sativum overexpressing VfAAP1 showed a significant increase [6]. Nitrogen supply during barley grain development is mainly dependent on the specific expression of HvAAP3 [33]. Moreover, the quantitative trait locus (QTL) qPC1, which controls rice grain protein content, was confirmed to be associated with the expression of OsAAP6 [7].

Bread wheat (Triticum aestivum L.) is one of the most important crops in the world, occupying 17% of cultivated land and accounting for approximately 35% of the world's staple food, and it is also an important protein source for humans [34, 35]. Genetically, wheat is an allohexaploid species with a complex origin and evolutionary history. The large and complex genome, over 17 Gbp in three homologous sub-genomes (A, B and D), poses a huge challenge for genome research in wheat. Recently, the release of the high-quality genome sequence of hexaploid wheat Chinese Spring (CS) based on the chromosomal strategy laid the foundation for the identification of wheat gene families at the genome level [36]. Although Wan et al. studied the temporal and spatial expression characteristics of the wheat AAT family and deduced their functions in nitrogen transport, the identification of the AAT family was incomplete and unsystematic, as it was based on the draft genome sequence and lacked an analysis of gene structure and evolutionary characteristics [37]. In this study, a genome-wide scan was conducted to identify the wheat AAT gene family. Then, the chromosome localization, gene structure, conserved protein motifs, duplication pattern, phylogenetic tree, and selective pressure of the putative wheat AAT genes were systematically analyzed. Finally, based on published transcriptome data and real-time PCR, the expression characteristics of these genes under different tissues/organs and different abiotic stresses were analyzed. Our work once again emphasizes the positive role of gene duplication and selection pressure in the expansion and functional diversification of the AAT gene family. Combining these genes’ temporal and spatial expression characteristics and their responses to abiotic stresses, this study will provide a basis for further functional analysis of the wheat AAT genes, as well as a better understanding of the molecular mechanisms underlying the wheat AAT genes’ regulation of wheat growth and stress responses.

Results

Identification of AAT gene family in wheat

Initially, a total of 307 putative wheat AAT transcripts were identified in the wheat genome by local BLASTP and HMMER searches. Twenty of the transcripts corresponded to 10 genes, and we selected the longest transcripts as candidates. One transcript containing incomplete conserved domains was omitted. Finally, 296 putative AAT genes with high confidence were identified in the wheat genome. Compared to other plant species reported, wheat had the largest AAT gene family, which might be the result of its allohexaploid genome and complex evolutionary process. Similar methods were used to identify the AAT family in maize and Brachypodium, with 107 AAT genes in maize and 80 in Brachypodium (Table 1). Based on subfamily classification and chromosomal localization, these 296 AATs in wheat were renamed, and detailed information, including gene structure and protein properties, is listed in supplementary
Table S1. The length of the putative wheat AAT proteins ranged from 318 to 1002, with isoelectric points (pI) ranging from 5.00 to 8.96 and molecular weights (Mw) ranging from 33.5 to 52.5 kD. The homologous AAT genes from different wheat sub-genomes showed no significant difference in protein characteristics.

To better understand the transmembrane structure of the AAT family, TMHMM Server 2.0 was used to predict the putative transmembrane (TM) regions. The number of TM regions in TaAAT proteins ranged from 6 to 15, and AAT genes belonging to the same subfamily showed a similar distribution (Fig. 1, Additional file 2: Table S1). Among them, CAT subfamily members contained the most TM regions, ranging from 13 to 15, while the AAP and LHT subfamily members contained the fewest TM regions, ranging from 7 to 11 and 7 to 10, respectively (Fig. 1). In addition, it was found that 19 groups of homologous genes derived from the A, B and D sub-genomes were mutated in TM number, accounting for approximately 20% (19 of 93) of the total complete homologous gene groups identified in this study, which confirmed that during the process of wheat genome doubling, the homologous genes from different sub-genomes had great TM number variation, to achieve functional synergy and suitability.

Chromosomal distribution and duplication analysis of AAT genes in wheat

To study the relationship between AAT gene family expansion and gene duplication in wheat, 294 of 296 TaAATs were mapped onto 21 chromosomes of wheat, and the remaining two genes were mapped onto the unattributed scaffold (Fig. 2). Except for the chromosomal inversion of chromosomes 4A and 4B, the number and distribution of TaAATs from different sub-genomes showed great similarity, confirming that the AAT family has been relatively conserved after hexaploid wheat formed. Moreover, wheat AAT family members showed significant heterogeneity in chromosome distribution, with relatively high density in specific chromosomal regions, such as the end of chromosome groups two and three, whereas the top of chromosome group six contained very few TaAATs. In terms of their overall distribution on chromosome groups, chromosome group two contained the most (61) AAT genes, while chromosome group one had only 20, and the remaining chromosome groups had 31 to 56 (Fig. 2).

Gene duplication is generally considered to be the major factor leading to gene family expansion and functional diversification. After two naturally interspecific hybridization events of three diploid species and doubling, current hexaploid wheat was produced. In principle, each wheat gene usually has three homologous loci caused by polyploidization [38]. Through sequence similarity analysis and chromosome location analysis, 93 homologous pairs containing three copies from the A, B, and D sub-genomes were found in the wheat genome, accounting for approximately 94% of all putative TaAATs. Although TaLAT3, TaATLb2 and TaAAP15 did not match our expectations of their chromosomal locations, the three sets were still considered homologs based on their high sequence similarity. Moreover, six AAT gene pairs contained only two copies among the A, B, or D homologous chromosomes, accounting for approximately 4% of all putative TaAATs, while five AAT genes contained only one copy. Only 8% of the AAT genes were lost, and the genes with the TM region mutation reached 20%, confirming that the functional adaptation of the wheat AAT family during wheat polyploidization was driven mainly by sequence mutations rather than copy number variation.

Among the wheat 296 TaAAT genes, 25.33% (75 of 296) came from tandem duplication events (Fig. 2). The 75 tandemly duplicated genes were divided into 32 groups, of which 24 groups contained 2 genes each, 5 groups contained 3 genes each
and 3 groups contained 4 genes each. Except for the two sets of tandemly duplicated genes only in the A and B sub-genomes, all of the tandemly duplicated genes had homologous copies in all three sub-genomes, indicating that most tandem duplication events occurred before wheat polyploidization. To assess the effect of tandem duplication events on the expansion of AAT subfamilies, we compared the proportions of tandemly duplicated genes in wheat, rice, *Arabidopsis* and potato subfamilies (Table 2). In all species, the proportion of tandemly duplicated genes in the AAP subfamily was very high, ranging from 25% to 52.63%, which confirmed the extensive contribution of tandem duplication events to the expansion of the AAP subfamily. In addition, for other subfamilies, dicotyledons and monocotyledons showed significant differences. For example, a high proportion of genes were tandemly duplicated in the ATLb subfamily in monocots, while the ProT subfamily, the ATLa subfamily, and the CAT subfamily in dicots had high tandemly duplicated gene ratios.

**Selective pressure analysis of AAT genes in wheat**

Gene duplication and natural selection largely shape the diversity of gene functions. In view of the tremendous impact of tandem duplication events on the expansion of the AAT family in wheat, we used the Ka/Ks value to evaluate the selection pressure of tandemly duplicated gene pairs (Fig. 3). Fifty tandemly duplicated gene pairs distributed in the same homologous group were selected for Ka/Ks analysis. The average value of all gene pairs was 0.27, which indicated that all genes were under purifying selection to maintain important biological roles. It is noteworthy that the median value and dispersion of Ka/Ks in wheat AAP subfamilies were significantly higher than those in other subfamilies, which indicated that a considerable number of *TaAAPs* have evolved new features under weak selection pressure with gene replication and sequence diversity (Fig. 3). All the above results indicate that chromosome doubling and tandem duplication were the key to the expansion of the wheat AAT family. In addition, tandem duplication events played an important role in the functional differentiation of the wheat AAP subfamily.

**Structural characteristics of 296 AAT genes in wheat**

Numerous studies have confirmed that one of the representative traces of family evolution is gene structural characteristics [39-41]. Based on the annotated genome structure information, we investigated all 296 wheat AAT genes (Fig. 4). Homologous genes derived from different sub-genomes exhibited similar intron/exon distributions in gene structure, suggesting that the AAT homologous genes in various sub-genomes of wheat are extremely functionally conserved. It is worth noting that the paralogous genes of the same subfamily showed some differences in gene structure, indicating that there was also a very significant functional differentiation among members of the same subfamily. For example, six of the 13 tandemly duplicated gene pairs in the AAP subfamily showed significant structural variation, which was consistent with previous Ka/Ks results, again confirming that the production of new functions between tandemly duplicated genes was the key to the functional diversification of the wheat AAP subfamily. Twenty-nine AAT genes did not contain introns, while the rest contained 1-16 introns, which was consistent with the studies in other species. In conclusion, these results suggest that to a certain extent, the structural characteristics of genes may determine the conservation or diversity of functions during family evolution.
To better explore the evolutionary relationships among AAT members of different subfamilies, conserved motifs were also predicted by MEME (Additional file: Figure S1). Similar to the gene structure, the distribution of conserved motifs in different gene families or subfamilies also showed some conservatism and variation. For example, Motif 1 was found in both the AAAP and APC families (although it was incomplete in APC family members), which confirmed that Motif 1 was relatively conserved in the expansion of the AAT family and was retained in both families. In contrast, Motifs 2 and 9 were only found in members of the AAAP family, and Motif 10 was mainly found in members of the APC family, suggesting that functional differentiation of the AAT gene family was accompanied by loss and variation of conserved motifs. In addition, some motifs were found to be distributed only in a few subfamilies. For example, Motif 3 was specific to the AAT subfamily, Motif 12 was specific to the ACT and CAT subfamilies, and Motif 8 was specific to the AAT and ANT subfamilies. Members of different families or subfamilies showed large differences, while members of the same subfamily shared similar conserved motifs. These results confirm that changes in the conserved sequences of AAT proteins played important roles in the functional differentiation of the AAT family.

Phylogenetic analysis and multiple-sequence alignment
To better reveal the subfamily classification of the AAT family in wheat, a phylogenetic tree including 707 protein sequences of AAT family members from rice, maize, Brachypodium, Arabidopsis thaliana, soybean and wheat AATs was constructed using the ML method (Fig. 5). The ML tree showed that all AAT proteins could be clearly divided into 12 independent branches with high confidence. The AAAP family contained 204 AAT proteins in wheat, while the APC family contained 92 AAT proteins. The AAAP family consisted of eight distinct subfamilies: amino acid permeases (AAPs, 66), lysine/histidine transporters (LHTs, 24), GABA transporters (GATs, 14), proline transporters (ProTs, 9), auxin transporters (AUXs, 15), amino acid transporter-like a (ATLα, 18), aromatic and neutral amino acid transporters (ANTs, 18) and amino acid transporter-like b (ATLβ, 40). The APC family consisted of four distinct subfamilies: the cationic amino acid transporters (CATs, 31), the amino acid/choline transporters (ACTs, 21), the polyamine H+-symporters (PHSs, 31) and tyrosine-specific transporters (TTPs, 9). A given subfamily’s genes from monocotyledonous and dicotyledonous plants were distributed in the same branch, confirming that the main features of the AAT family were formed before the differentiation of monocotyledonous and dicotyledonous plants.

The alignment of the TaANT members is shown in Fig. 6 as an example. All ANT family member protein sequences derived from the wheat A sub-genome were used for multiple sequence alignments. The overall identity of the protein sequences of these genes was 61.97%. There were six conserved motifs in TaANTs, including motifs 1, 7, 9, 2, 13 and 5 (Fig. 6). There was a very high correlation between the conserved motifs and the transmembrane regions, in which motifs 1, 7, 2, 5, 13 corresponded to TM1, TM2, TM7, TM9 and TM8, respectively, and motif 9 corresponded to TM3. All of the conserved sequences were located around the transmembrane domains, suggesting that the stabilization of the transmembrane domains played an integral role in the normal functioning of ANT family proteins. The unequal distribution of TM numbers in different subfamilies also confirmed that TM affected the functional specificity of different subfamilies.
Analysis of cis-regulatory elements and three-dimensional modeling

The cis-regulatory elements predicted in promoter regions of all AAT genes were mainly classified into three categories: tissue specificity, stress response, and hormone response. A large number of stress response cis-regulatory regulatory elements were found in all gene promoter regions, indicating that the AAT genes responded strongly to stress, and multiple GA, ABA and other hormone response elements were found, indicating that these genes might be involved in multiple hormone signal pathways. In addition, multiple transcription factor binding sites, such as MYB, were observed in the promoter regions, confirming that these genes might be regulated by a variety of transcription factors.

All predicted AAT proteins contained multiple α-helices and coil structures (Fig. 7). The multiple α-helix structures ensured the efficient and stable transmembrane transport of AAT proteins. Most AAT proteins had similar three-dimensional structures, and the closer the phylogenetic relationship was between genes, the closer the three-dimensional structures of the proteins were, such as LHT and ProT (Fig. 7). This indicates that the maintenance of the three-dimensional structure plays an important role in the function of AAT family proteins.

Spatiotemporal expression patterns of AATs in wheat

To determine the expression pattern of the TaAAT genes, we downloaded the transcriptome data from different tissues at different growth stages from the public expression database. The gene pairs containing three copies accounted for 94% of the total AAT genes, so for convenience of display, the average TPM value of the homologous gene pair expression was used for expression level analysis. The log2(TPM+1) value was used for the heat map display (Fig. 8). The results showed that the wheat AAT genes had many different expression patterns. Some genes of the AAP, ATLa, CAT, and AUX subfamilies were highly expressed in multiple tissues, such as TaAAP1, TaAAP14, TaATLa4, TaATLa5, TaAUX3, TaCAT6 and TaCAT11, while members of the remaining families exhibited high expression of specific tissues or specific organs. For example, TaANT3 was specifically expressed in peduncle, while TaTTP1 was specifically expressed in leaf. In addition, the same gene showed differential expression at different developmental stages. For example, TaBAT4 was specifically expressed only in leaves during the grain-filling stage and not at the seedling or flag leaf stage.

Gene duplication and expression patterns of duplicated AAT genes

Gene duplication events can serve as a key mechanism for increasing gene family diversity, particularly through nonfunctionalization, subfunctionalization, and new functionalization of duplicated genes. Subfunctions and new functional duplication lead to functional diversity within the family and can be expressed in different tissues or at different developmental stages than their progenitor cells [43]. Gene replication and diversification events are well documented in Arabidopsis [44]. We observed that most of the repeated AAT genes in wheat were differentially expressed in the tissue/organ/developmental stage (Fig. 9). Based on the gene expression pattern, we observed three functional variations in homologous gene pairs in wheat. For example, we observed nonfunctionalization in the TaAAP8/TaAAP9 gene pair. Specifically, TaAAP9s were expressed in root, peduncle and shoot apical meristem, while TaAAP8s were almost unexpressed in all tissues. TaAAP21s were expressed in all tissues, while TaAAP22s were not. Subfunctional phenomena were also observed, such as the expression levels of TaLHT7s being significantly weaker than those of TaLHT8s, and
similar phenomena were also observed in the TaAAP16/TaAAP17 pair. In addition, gene duplication has produced new functions. Collectively, our results show that gene duplication events play integral roles in the generation of new functions and the retention of key functions during species evolution. In addition, the expression profiles of repeated wheat AAT proteins showed that most of them had undergone subfunctionalization. These observations are consistent with those in other plant species, in which closely related genes have different expression patterns.

**Response of TaAATs to abiotic stresses**

To explore the response of the amino acid transporter family genes to abiotic stresses, three major abiotic stresses, drought, heat, and salt, were simulated, and 12 TaAATs highly expressed in leaves or roots were selected for qRT-PCR (quantitative real time polymerase chain reaction) analysis (Fig. 10). Both TaAATs highly expressed in leaves and TaAATs highly expressed in roots showed different response patterns to different abiotic stresses. The expression of TaAAP2 in leaves exhibited rapid down-regulation under drought stress but decreased first and then increased with increasing heat stress time. The expression levels of TaAAP3 and TaATL3 increased slightly at 1 h of drought and heat stresses and increased significantly at 6 h. Unlike these genes, TaLHT8 was significantly up-regulated at the early stage of drought stress (1 h) but inhibited at 6 h. In addition, TaATLa2 showed a sustained response to drought stress.

All six selected genes showed different degrees of response to salt stress, and their expression patterns could be roughly divided into two categories, up-regulation and down-regulation. TaAAP7, TaAAP17, TaAAP18 and TaLHT3 were up-regulated under salt stress. TaAAP7 and TaLHT3 maintained high expression at 48 h after salt stress, while TaAAP17 and TaAAP18 were down-regulated, which confirmed that the response of TaAATs to salt stress was similar to that of drought and heat stress, showing different response intensities and durations to improve the adaptability of wheat to salt stress. Moreover, with increasing salt stress time, the expression levels of TaANT5 and TaBAT2 decreased continuously, suggesting that salt stress might seriously affect the functions of these two genes.

**Discussion**

**AAT gene family in wheat**

Plant amino acid transporters play important roles in processes of seed germination, seedling growth, grain quality formation and response to pathogens and abiotic stresses by adjusting the transport and distribution of different amino acids [3-5, 12]. Recently, AAT gene family has been systematically identified and characterized in multiple species [15-17, 29]. However, due to the complex genome of wheat, it has not been thoroughly and systematically characterized in wheat. In this study, we identified and characterized AAT gene family in wheat through genome-wide analyses, and studied its evolutionary model, tissue expression patterns and response to abiotic stress.

The number of reported AAT genes varied across various high plants, and ranged from 63 to 189, with 63 in Arabidopsis [29], 72 in potato [17], 189 in soybean [16]...
and 85 in rice [15]. In the present study, 297, 107 and 80 AAT genes were identified in wheat, maize and Brachypodium (Table 1). The difference of AAT family members between monocots and eudicots was mainly due to the unequal expansion of different subfamilies. For example, the number of AAP subfamily members in monocots was significantly greater than that of eudicots, which indicated that the expansion rate of AAP subfamily was different after monocots and eudicots differentiation. The varied numbers of AAT genes in different species may be due to the gene duplication events, including tandem, segmental or whole-genome duplication (WGD) events. In fact, 279 of 297 identified AAT genes formed 93 homologous pairs containing three copies from A, B and D sub-genome, accounting for 94% of all putative AAT genes, confirming that WGD event was the main driver of AAT family expansion in wheat. Secondly, 25.33% (75 of 296) were related to tandem duplication events, and 30 of 32 tandemly duplicated gene groups had homologous copies in all three sub-genomes, which suggested that most occurred before the formation of hexaploid wheat (Fig. 2). In general, the expansion of AAT gene family in wheat mainly relies on WGD and tandem duplication events.

**Functional diversity of wheat AAT family members**

To improve the adaptability to the environment, the evolution of plants is usually accompanied by the generation of a large number of new genes and subsequent functional differentiation. Gene functional differentiation mainly includes three levels of variation: sequence variation, structural variation, and expression level variation. Natural selection is a key determinant of gene functional diversification. The ratio between the number of nonsynonymous substitutions per nonsynonymous site (Ka) and the number of synonymous substitution per synonymous site (Ks) reflects the selection pressure of duplicated genes. We used Ka/Ks values to evaluate the effect of natural selection on functional differentiation (Fig. 3). Overall, the Ka/Ks values for all tandemly duplicated gene pairs evaluated were less than 1, confirming that these genes have undergone purification selection to maintain important biological functions. On the other hand, the median value and dispersion of the Ka/Ks values of wheat AAP subfamily members were much higher than those of other subfamily members, which indicated that the AAP subfamily was more likely to generate new features during the expansion process. This is consistent with the huge difference in the number of AAP subfamily members of monocots and eudicots (Table 1). In addition, about 20% (19 of 93) of the AAT gene homologous pairs from A, B and D sub-genomes produced variation in TM number, which was not due to the variation of conserved motifs to a large extent (Table 1, Additional file 2: Table S1). All of these confirmed that the AAT family in wheat improved the diversity, coordination and adaptability of AAT family members through the different sequence variation of different subfamilies and different sub-genomic homologous genes in the expansion process.

Previous studies have confirmed that one of the representative traces of gene family evolution is the variation of gene structure [39-41]. Gene structure from the same AAT subfamily showed some differences while maintaining conservation
(Figure 4). These differences were especially obvious in the members of AAP
subfamily. The number of introns of AAP genes ranged from 0-6, and there were 6
pairs of significant structural variation in 13 tandemly duplicated gene pairs, which
was consistent with our results about the difference distribution of AAP subfamily
members in monocots and eudicots, the large median value and the dispersion degree
of Ka/Ks. All of this confirmed that the expansion and functional differentiation of
AAP subfamily have a positive effect on the evolution of AAT family and the
improvement of wheat adaptability. In addition, the analysis of conservative motif of
AAT family members showed that some motifs existed conservatively in all
subfamilies, such as motif 1, while some motifs were unique to different subfamilies,
such as motif 2, motif 9 and motif 10 (Additional file: Figure S1). The former may
determine the important basic functions of AATs in wheat, while the latter may affect
some specific new functions. In short, changes in gene structure and deletion and
mutation of conserved motifs were also one of the important reasons for the
functional diversification of wheat AAT family members.

Variation in the expression level of newly duplicated genes is an important way
for functional differentiation after gene family expansion, including
nonfunctionalization, subfunctionalization, and new functionalization. The generation
of subfunctionalization and new functionalization makes the newly duplicated genes
different from the ancestral genes in terms of expression level, spatial-temporal
specificity of expression to perform different functions [42, 43]. We observed all three
types among the wheat AAT duplicated gene pairs, such as TaAAP8/TaAAP9 gene pair
showed significant nonfunctionalization of TaAAP8s, TaLHT7/TaLHT8 gene pair
showed subfunctionalization of TaLHT7s, and TaATLb12s and TaATLb13s showed
different spatiotemporal expression characteristics (Fig. 9). In addition, the expression
levels of homologous genes from different sub-genomes were not the same. For
example, TaBAT4-3A and TaBAT4-3B showed higher expression levels than
TaBAT4-3D in all tissues, while TaAAP16-6A and TaAAP16-6D had higher expression
levels than TaAAP16-6B. This indicated that during the formation of hexaploid wheat,
the expression level of AAT homologous genes from different sub-genomes changed,
thereby enhancing the overall coordination of gene pairs. Actually, the functional
differentiation of wheat AAT family is the result of the combined effects of sequence,
structure and expression levels of AAT family members, and the multi-level mutations
of AAT homologous genes from different sub-genomes have a very positive effect to
the functional adaptation of AAT genes in wheat.

Spatiotemporal expression patterns of TaAAT genes and response to abiotic stresses
The analysis of TaAAT genes on spatiotemporal expression patterns may provide
useful information for establishing their putative functions. TaAAT genes showed
different expression patterns at different developmental stages in different tissues, and
some genes are highly expressed throughout the whole growth process, such as
TaAAP1, TaAAP14, TaATLa4, TaAUX3, TaCAT6, indicating that these genes were
critical for the overall growth and development of plants. Some genes showed high
expression in specific tissues or organs, such as TaANT3 and TaTTP1. Some genes
are expressed at specific developmental stages, such as TaBAT4. The specific expression of these genes in specific tissues or organs at different stages indicated that these genes may perform specific functions in specific tissues at specific developmental stages.

AAP members in Arabidopsis play critical roles in nutrient transport during seed development or long-distance transport of amino acids [28, 29]. AtAAP1 participates in the uptake of amino acids into root cells [10]. We also found that TaAAP19, which was located in the same branch of AtAAP1 in the phylogenetic tree, was highly expressed in roots, suggesting that TaAAP19 and AtAAP1 might have similar functions. Moreover, TaAAP19, TaAAP7 and TaAAP17 also showed similar expression characteristics, suggesting that wheat improved its adaptability by increasing the fault tolerance of TaAAPS. AtAAP3 plays an important role in the uptake of amino acids in the xylem of Arabidopsis [42]. TaAAP1 was highly expressed in SC and SSAM, and considering its close relationship with AtAAP3, we speculated that TaAAP1 had similar functions in wheat. Grain protein content (GPC) is an important determinant of nutritional quality in cereals, and studies based on Arabidopsis and rice have confirmed that both AtAAP8 and OsAAP6 are related to GPC [28, 4]. We also found high expression of TaAAP8 in wheat grain, confirming that the functions of these genes were relatively conserved between monocots and dicots. In our study, in addition to the homologous genes identified in other species, we also observed the specific expression of several genes, such as TaANT5, TaAUX5, TaLHT2, and TaBAT2, in root tissues. Obviously, these genes contribute to the uptake and transport of amino acids in roots. Moreover, the high expression of TaAAP4, TaAAP20, TaAUX2, TaProT1, TaLAT5 and TaTTP3 in grains was indispensable for the accumulation of protein.

Drought, heat and salinity are the three main abiotic stresses facing crops, which may greatly reduce crop growth and productivity. The accumulation of osmotic active compounds in plants is an important way to balance osmotic potential under drought or salt stress, and the changes in amino acid composition in plants are mainly regulated by amino acid transporters [45]. It has been proven that in many species, excessive accumulation of proline could enhance the tolerance of plants to osmotic stress [28, 46, 47]. In this study, based on qRT-PCR analysis of AAT genes that were highly expressed in wheat under different abiotic stresses, we identified multiple AAT genes that responded specifically/non-specifically to different abiotic stresses, and these genes were found in Different abiotic stresses show different response patterns. Some genes, such as TaAAP3, TaATLa2, and TaATLb13, showed sustained responses under drought and high temperature stress. These genes may play important roles in maintaining normal amino acid transport in wheat under long-term drought and high temperature stress. TaAAP2 and TaLHT8 were up-regulated in the late stage of high temperature stress and the early stage of drought stress, respectively. The specific responses of these two genes to different stresses in different stages increased the drought and heat stress resistance. Similar to the response under high temperature and drought stress, under salt stress, different TaAAT genes also showed different response intensity and duration. TaAAP7 and TaLHT3 remained very high after 48 hours of salt stress, while TaAAP17 and TaAAP18 were down-regulated, confirming that the response of the first two genes to salt stress can be maintained for a longer period of time, while the latter two genes may play a role mainly in the early stage. In general, TaAATS showed different response modes to abiotic stresses, including enthusiasm for response, generality of response, and differences in response duration, which may greatly improve the adaptability of wheat to abiotic stresses.
Conclusions

We identified 296 AAT gene family members in the wheat genome. Similar to other identified species, all wheat AAT genes were classified into two families - the AAAP and APC families. The AAAP family was subdivided into 8 subfamilies, while the APC family was divided into 4 subfamilies. We demonstrate that the expansion of the wheat AAT gene family is primarily due to WGD and tandem duplication events, while tandem repeat events have greatly determined the functional differentiation of AAT family members. We systematically outlined the chromosomal distribution, gene structure, phylogeny and conserved motifs of AAT family members in wheat and annotated all its AAT genes. We further evaluated the expression patterns of wheat AAT gene family members in different tissues and their responses to three conventional abiotic stresses, heat, drought and salt stress. We also identified several important candidate genes that might affect grain quality and root amino acid transport. Our work will provide a comprehensive framework for the study of the wheat AAT family and will also contribute to the functional analysis and utilization of wheat AAT genes.

Methods

Identification of the AAT gene family in wheat

Several methods were used to discover putative AAT family members in the wheat genome. First, wheat whole-genome protein sequences (IWGSC RefSeq v1.1) were downloaded from the Wheat URGI database (https://wheat-urgi.versailles.inra.fr) to construct a local protein database. Then, the known protein sequences of 337 AAT genes from Arabidopsis, rice and soybean were used as seed sequences to search the wheat protein database through the local BLASTP program with an e-value of 1e-5 and identity of 50% as the threshold. Furthermore, the hidden Markov model (HMM) profiles of the AAT domain (PF00324 and PF01490) were downloaded from the PFAM database (http://pfam.xfam.org/), and all putative AAT proteins predicted by BLASTP were further screened by conserved domains using the HMMER search tool [48]. An NCBI conserved domain database (CDD) search was also used to check the conserved protein domains of these candidate genes (https://www.ncbi.nlm.nih.gov/cdd). After removing the redundant sequences, the remaining sequences were submitted to InterProScan (http://www.ebi.ac.uk/interpro/scan.html) to reconfirm the existence and integrity of the AAT conserved domains. Finally, all putative AAT gene family members in wheat were identified after removing the sequences that did not contain the entire conserved domains. Information about these putative wheat AAT genes (TaAATs), including full-length cDNA accessions, coding sequence length, and gene structure, was obtained from the genome annotation gff3 file. An online tool, Gene Structure Display Server GSDS (http://gsds.cbi.pku.edu.cn/), was used to determine the structures of these genes, and the Computer pl/MW tool in the ExPASy database (https://web.expasy.org/compute_pi/) was used to calculate the biochemical parameters, including the theoretical pi (isoelectric point) and Mw (molecular weight) [49]. Moreover, the TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) was applied with default settings to predict the putative transmembrane (TM) regions in each TaAAT protein.

Chromosomal localization, duplication and selective pressure of AAT genes in wheat

TaAAT genes were mapped onto the chromosomes by identifying their chromosomal positions obtained from the wheat genome annotation gff3 file. The Multiple
Collinearity Scan toolkit (MCScanX) program was used to investigate gene duplication, and manual screening was performed according to the mature method described by Wang et al. [50, 51]. A TBtools tool was used to visualize the chromosome localizations and duplicated regions of all TaAAT genes [52].

Natural selection shapes the diverse functions of duplicated genes. To assess the impact of sequence duplication on function, KaKs_Calculator 2.0 was used to calculate the nonsynonymous (Ka) and synonymous (Ks) rates (Ka/Ks) of each aligned codon in the pairs of duplicated AAT genes [53].

**Phylogenetic analysis and multiple sequence alignment**

To clarify the evolutionary relationship between AAT gene family members in wheat and other angiosperms and to classify their subfamilies, AAT gene family members of three monocotyledonous species, maize, rice, Brachypodium, and two dicotyledonous species, soybean and Arabidopsis, were compared using a phylogenetic tree together with TaAAT genes. Multiple-sequence alignment of the protein sequences was performed using MUSCLE 3.8 [54]. The phylogenetic tree was constructed using the maximum likelihood (ML) method by PhyML 3.1 (http://www.atgc-montpellier.fr/phyml/versions.php) with the JTT model selected from IQ-TREE [55]. iTOL v3 (http://itol.embl.de/#) was used to display the phylogenetic tree.

The Multiple EM for Motif Elicitation (MEME) program was used to determine the conserved protein motifs of these genes, and the parameters were as follows: the optimal motif width was between 6 and 200 residues, allowing the presence of any number of repeating motif sites, and the maximum number of motifs was 20 [56]. The amino acid sequence conservation of TaANT subfamily members was analyzed using DNAMAN software, and the conserved motifs and TM regions were manually annotated.

**Identification of cis-regulatory elements and prediction of three-dimensional modeling**

The sequence 2000 bp upstream of the start codon was considered the promoter region for each gene, and the promoter sequences were extracted from each genome using the SAMtools program [57]. The putative transcriptional response elements of these gene promoters were predicted using the PlantCARE server (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/), a database of plant cis-acting regulatory elements [58].

To determine the differences in the structure of different AAT subfamily proteins and their effects on functions, the three-dimensional structure of a representative AAT protein from each subfamily was determined using the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2) [59].

**Spatiotemporal expression patterns of AATs in wheat**

To explore the spatiotemporal expression patterns of TaAATs, published wheat transcriptome data were downloaded from expVIP (http://www.wheat-expression.com) [60]. These data include read count values from 14 tissues at three important stages of wheat growth and development, including the seedling stage, flag leaf stage and milk grain stage. The tissues used at the seedling stage included radicle (SRA), coleoptile (SC), stem axis (SSA), first leaf blade (SFL), roots (SR) and shoot apical meristem (SSAM); the tissues at the flag leaf stage included flag leaf blade (FFL), shoot axis (FSA), roots (FR); and the tissues at the milk grain stage included flag leaf blade (MFL),
peduncle (MP), awns (MA), glumes (MGL), and grain (MG). The normalized expression levels of *TaAATs* were expressed by calculating TPM (transcripts per million) values standardized using the R package edgeR [61]. The log2(TPM+1) values are displayed in a heat map to show the tissue expression characteristics of *TaAATs*.

### Plant materials, abiotic stress treatments and quantitative real-time PCR

For stress treatments, the wheat seeds (cultivar ‘Chinese Spring’) were collected from the experimental field of our own laboratory in College of Agronomy of Northwest A&F University. Chinese Spring is a local wheat cultivar and have been kept in our laboratory for more than thirty years. All seeds planted in a greenhouse of Northwest A&F University with a temperature of 22°C, while a 16 h photoperiod (12000 lux) and an 8 h dark period were used for seed growing. Seedlings at the three-leaf stage of wheat were used for the abiotic stress treatments. Seedlings were exposed to 20% polyethylene glycol (PEG-6000) and 35°C temperature for 1 h and 6 h to simulate drought and heat stress, respectively [17]. In addition, long-term soaking of seedlings in NaCl solution (200 mM; 6 h, 12 h, 24 h, 48 h) was used to simulate salt stress [62]. Then, the plant leaves and roots were collected for RNA extraction with an RNAiso Plus Kit (Takara, China). A PrimeScriptTM II 1st Strand cDNA Synthesis Kit (Takara, China) and a SuperReal PreMix Color (SYBR Green) Kit were used for the synthesis of the first strand complementary cDNAs and the quantitative real time-PCR, respectively. TaActin was used as an internal control, and all the special primers for qRT-PCR were designed using Primer Premier 6.0 software (http://www.premierbiosoft.com/primerdesign/). All primers are listed in Additional file 3: Table S2. These primers cannot distinguish three homologous copies, so the expression level of the selected gene would be amplified. Each sample contained three biological replicates, and the $2^{-\Delta\Delta C_{t}}$ method was used to evaluate the expression level.

### Abbreviations

- **AAT**: Amino acid transporter; **APC**: Amino acid-polyamine-choline; **AAAP**: amino acid/auxin permease; **ACT**: Amino acid/choline transporter; **CATs**: Cationic amino acid transporter; **PHS**: Polyamine H+-symporter; **TTP**: tyrosine-specific transporter; **LHT**: Lysine and histidine transporter; **GAT**: γ-aminobutyric acid transporter; **ProT**: Proline transporter; **AAP**: amino acid permease; **AUX**: auxin transporter; **ANT**: aromatic and neutral amino acid transporter; **GABA**: γ-aminobutyric acid; **QTL**: Quantitative trait locus; **pI**: isoelectric point; **Mw**: molecular weight; **TM**: transmembrane; **WGD**: whole-genome duplication; **qRT-PCR**: quantitative real time polymerase chain reaction.

### Acknowledgements

Not Applicable

### Authors' contributions

RT and YY analyzed the data. RT, YY and MC designed the work and wrote the manuscript. All authors have read and approved the manuscript.

### Funding

This work was funded by the National Natural Science Foundation of China (no. 31772160). The funding supported the study in the chemicals purchase and publication fee.
Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ortiz-Lopez A, Chang H, Bush DR. Amino acid transporters in plants. Biochim Biophys Acta. 2000;1465(1-2):275-80.
2. Singh BK. Plant Amino Acids: Biochemistry and Biotechnology. Marcel Dekker, New York. 1999;227-47.
3. Williams LE, Bick JA, Neelam A, Weston KN, Hall JL. Biochemical and molecular characterization of sucrose and amino acid carriers in Ricinus communis. J Exp Bot. 1996;47:1211-16.
4. Williams LE, Miller A. Transporters responsible for the uptake and partitioning of Nitrogenous Solutes. Annu Rev Plant Physiol Plant Mol Biol. 2001;52:659-88.
5. Miranda M, Borisjuk L, Tewes A, Heim U, Sauer N, Wobus U, Weber H. Amino acid permeases in developing seeds of Vicia faba L.: expression precedes storage protein synthesis and is regulated by amino acid supply. Plant J. 2001;28(1):61-71.
6. Rolletschek H, Hosein F, Miranda M, Heim U, Gotz KP, Schlereth A, Borisjuk L, Saalbach I, Wobus U, Weber H. Ectopic expression of an amino acid transporter (VfAAP1) in seeds of Vicia narbonensis and pea increases storage proteins. Plant Physiol. 2005;137(4):1236-49.
7. Peng B, Kong HL, Li YB, Wang LQ, Zhong M, Sun L, Gao GJ, Zhang QL, Luo LJ, Wang GW, et al. OsAAP6 functions as an important regulator of grain protein content and nutritional quality in rice. Nat Commun. 2014;5:4847.
8. Tegeder M, Ward JM. Molecular evolution of plant AAP and LHT amino acid transporters. Front Plant Sci. 2012;3:21.
9. Koch W, Kwart M, Laubner M, Heineke D, Stransky H, Frommer WB, Tegeder M. Reduced amino acid content in transgenic potato tubers due to antisense inhibition of the leaf H+/amino acid symporter StAAP1. Plant J. 2003;33(2):211-20.
10. Lee YH, Foster J, Chen J, Voll LM, Weber APM, Tegeder M. AAP1 transports uncharged amino acids into roots of Arabidopsis. Plant J. 2007;50(2):305-19.
11. Svennerstam H, Ganeteg U, Nasholm T. Root uptake of cationic amino acids by Arabidopsis depends on functional expression of amino acid permease 5. New Phytol. 2008;180(3):620-30.
12. Liu X, Bush DR. Expression and transcriptional regulation of amino acid transporters in plants. Amino Acids. 2006;30(2):113-20.
13. Lehmann S, Gumy C, Blatter E, Boeffel S, Fricke W, Rentsch D. In planta function of compatible solute transporters of the AtProT family. J Exp Bot. 2011;62(2):787-96.
14. Rentsch D, Schmidt S, Tegeder M. Transporters for uptake and allocation of organic nitrogen compounds in plants. FEBS Lett. 2007;581(12):2281-9.

15. Zhao H, Ma H, Yu L, Wang X, Zhao J. Genome-wide survey and expression analysis of amino acid transporter gene family in rice (Oryza sativa L.). PLoS ONE. 2012;7(11):e49210.

16. Cheng L, Yuan HY, Ren R, Zhao SQ, Han YP, Zhou QY, Ke DX, Wang YX, Wang L. Genome-wide identification, classification, and expression analysis of amino acid transporter gene family in glycine max. Front Plant Sci. 2016;7:515.

17. Ma H, Cao X, Shi S, Li S, Gao J, Ma Y, Zhao Q, Chen Q. Genome-wide survey and expression analysis of the amino acid transporter superfamily in potato (Solanum tuberosum L.). Plant Physiol Biochem. 2016;107:164-77.

18. Okumoto S, Pilot G. Amino acid export in plants: a missing link in nitrogen cycling. Mol Plant. 2011;4(3):453-63.

19. Fischer WN, Andre B, Rentsch D, Krolkiewicz S, Tegeder M, Breitkreuz K, Frommer WB. Amino acid transport in plants. Trends Plant Sci. 1998;3(5):188-95.

20. Hunt E, Gattolin S, Newbury HJ, Bale JS, Tseng HM, Barrett DA, Pritehard J. A mutation in amino acid permease AAP6 reduces the amino acid content of the Arabidopsis sieve elements but leaves aphid herbivores unaffected. J Exp Bot. 2012;61(1):55-64.

21. Saier MH, Reddy VS, Tamang DG, Vastermark A. The transporter classification database. Nucleic Acids Res. 2014;42(D1):D251-8.

22. Serrano R. Salt tolerance in plants and microorganisms: toxicity targets and defense responses. Int Rev Cytol. 1996;165:1-52.

23. Lehmann S, Funck D, Szabados L, Rentsch D. Proline metabolism and transport in plant development. Amino Acids. 2010;39(4):949-62.

24. Schwacke R, Grallath S, Breitkreuz KE, Stransky E, Stransky H, Frommer WB, Rentsch D. LeProT1, a transporter for proline, glycine betaine, and γ-amino butyric acid in tomato pollen. Plant Cell. 1999;11(3):377-91.

25. Ueda A, Shi W, Shimada T, Miyake H, Takabe T. Altered expression of barley proline transporter causes different growth responses in Arabidopsis. Planta. 2008;227(2):277-86.

26. Marchant A, Bhalarao R, Casimiro I, Eklof J, Casero PJ, Bennett M, Sandberg G. AUX1 promotes lateral root formation by facilitating indole-3-acetic acid distribution between sink and source tissues in the Arabidopsis seedling. Plant Cell. 2002;14(3):589-97.

27. Marchant A, Kargul J, May ST, Muller P, Delbarre A, Perrot-Rechenmann C, Bennett MJ. AUX1 regulates root gravitropism in Arabidopsis by facilitating auxin uptake within root apical tissues. Embo J. 1999;18(8):2066-73.

28. Okumoto S, Schmidt R, Tegeder M, Fischer WN, Rentsch D, Frommer WB, Koch W. High affinity amino acid transporters specifically expressed in xylem parenchyma and developing seeds of Arabidopsis. J Biol Chem. 2002;277(47):45338-46.

29. Su YH, Frommer WB, Ludewig U. Molecular and functional characterization of a family of amino acid transporters from Arabidopsis. Plant Physiol. 2004;136(2):3104-13.

30. Sanders A, Collier R, Trehewey A, Gould G, Sieker R, Tegeder M. AAP1 regulates import of amino acids into developing Arabidopsis embryos. Plant J. 2009;59(4):540-52.

31. Zhang L, Tan Q, Lee R, Trehewey A, Lee YH, Tegeder M. Altered xylem-phloem transfer of amino acids affects metabolism and leads to increased seed yield and oil content in Arabidopsis. Plant Cell. 2010;22(11):3603-20.
32. Schmidt R, Stransky H, Koch W. The amino acid permease AAP8 is important for early seed development in *Arabidopsis thaliana*. Planta. 2007;226(4):805-13.

33. Kohl S, Hollmann J, Blatter FR, Radchuk V, Andersch F, Steuernagel B, Schmutzer T, Scholz U, Krupinska K, Weber H, et al. A putative role for amino acid permeases in sink-source communication of barley tissues uncovered by RNA-seq. BMC Plant Biol. 2012;12(1):154.

34. Gill BS, Appels R, Botha-Oberholster AM, Buell CR, Bennetzen JL, Chalhoub B, Chumley F, Dvorak J, Iwanaga M, Keller B, et al. A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. Genetics. 2004;168(2):1087-96.

35. International Wheat Genome Sequencing Consortium (IWGSC). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science. 2014;345(6194):1251788.

36. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. 2018;361(6403):eaar7191.

37. Wan YF, King R, Mitchell RAC, Hassani-Pak K, Hawkesford MJ. Spatiotemporal expression patterns of wheat amino acid transporters reveal their putative roles in nitrogen transport and responses to abiotic stress. Sci Rep. 2017;7(1):5461.

38. Feldman M, Levy AA. Allopolyploidy - a shaping force in the evolution of wheat genomes. Cytogenet Genome Res. 2005;109(1-3):250-8.

39. Javelle M, Klein-Cosson C, Vernoud V, Boltz V, Maher C, Timmermans M, Deppege-Fargeix N, Rogowsky PM. Genome-wide characterization of the HD-ZIP IV transcription factor family in maize: preferential expression in the epidermis. Plant Physiol. 2011;157(2):790-803.

40. Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, Tang YX. Genome-wide analysis of the MYB transcription factor superfamily in soybean. BMC Plant Biol. 2012;12(1):106.

41. Hudson KA, Hudson ME. A classification of basic helix-loop-helix transcription factors of soybean. Int J Genomics. 2015;10.

42. Okumoto S, Koch W, Tegeder M, Fischer WN, Biehl A, Leister D, Stierhof YD, Frommer WB. Root phloem-specific expression of the plasma membrane amino acid proton co-transporter AAP3. J Exp Bot. 2004;55(406):2155-68.

43. Bhattcharjee A, Ghangular R, Garg R, Jain M. Genome-wide analysis of homeobox gene family in legumes: identification, gene duplication and expression profiling. PLoS ONE. 2015;10(3):e0119198.

44. Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. Mol Biol Evol. 2005;23(2):469-78.

45. Tarczynski MC, Jensen RG, Bohnert HJ. Stress protection of transgenic tobacco by production of the osmolyte mannitol. Science. 1993;259(5094):508-10.

46. Rentsch D, Hirner B, Schmelzer E, Frommer WB. Salt stress-induced proline transporters and salt stress-repressed broad specificity amino acid permeases identified by suppression of a yeast amino acid permease-targeting mutant. Plant Cell. 1996;8(8):1437-46.

47. Wu M, Wu SN, Chen Z, Dong Q, Yan HW, Xiang Y. Genome-wide survey and expression analysis of the amino acid transporter gene family in poplar. Tree Genet Genomes.
Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics. 2013;29(19):2487-9.

Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The Proteomics Protocols Handbook. 2005;571-607.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.

Wang M, Yue H, Feng K, Deng P, Song W, Nie X. Genome-wide identification, phylogeny and expression profiles of mitogen activated protein kinase kinase kinase (MAPKKK) gene family in bread wheat (Triticum aestivum L.). BMC Genomics. 2016;17(1):668.

Chen C, Xia R, Chen H, He Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. bioRxiv. 2018;289660.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinf. 2010;8(1):77-80.

Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5(1):113.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2014;32(1):268-74.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:202-8.

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987-93.

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombouts S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325-7.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015;10:845-58.

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, et al. The transcriptional landscape of polyploid wheat. Science. 2018;361(6403):eaar6089.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.

Wu GQ, Wang JL, Li SJ. Genome-wide identification of Na+/H+ Antiporter (NHX) Genes in Sugar Beet (Beta vulgaris L.) and their regulated expression under salt stress. Genes. 2019;10:401.
Figure legends

Fig. 1. The distribution of the putative transmembrane (TM) region numbers in the wheat AAT family. The X-axis lists the 12 AAT subfamilies in wheat, and the Y-axis represents the range of TM region numbers in each subfamily. In the boxplot, the middle line indicates the median, and the box indicates the range from the 25 to the 75th percentile of the total TM number. The top and bottom points represent the maximum and minimum values, respectively.

Fig. 2. Chromosomal localization and gene duplication events of TaAAT genes. Chromosome numbers are shown on the left side of each strip. Homologous pairs are linked by red lines. Blue shadows indicate tandem duplication.

Fig. 3. Effects of tandem duplication events on the functional differentiation of wheat AAT gene subfamilies. The Y-axis indicates the rates of the nonsynonymous (Ka) and synonymous (Ks) substitutions (Ka/Ks). The X-axis shows the subfamilies containing tandemly duplicated AAT genes.

Fig. 4. Gene structure of TaAAT genes in each subfamily. Green boxes represent exons, and the block line represents introns. The UTRs are indicated by yellow boxes. The sizes of introns and exons can be estimated using the scale at the bottom.

Fig. 5. Phylogenetic tree of AAT proteins in wheat, maize, Brachypodium, Arabidopsis and soybean. Multiple sequence alignment was performed by MUSCLE, and the phylogenetic tree was constructed using PhyML with the maximum likelihood (ML) method.

Fig. 6. Multiple-sequence alignment of the transmembrane (TM) region of the ANT subfamily in wheat. The blue line represents the TM regions of TaANT proteins, and the red box represents the conserved motifs predicted by MEME. Identical (100%), conserved (75–99%), and blocks of similar (50–74%) amino acid residues are shaded in red, blue, and yellow, respectively.

Fig. 7. Three-dimensional modeling of AAT proteins in wheat. AAT proteins were selected from 12 subfamilies for three-dimensional structure prediction and display, with a confidence level > 90%. The displayed AAT proteins are TaAAP2-2A, TaGAT1-1A, TaLHT1-1A, TaProT1-2A, TaAUX1-1A, TaATLa1-3A, TaANT1-2A, TaATLb9-4A, TaCAT2-2A, TaLAT1-2A, TaBAT1-2A and TaTTP1-2A.

Fig. 8. The expression levels of wheat AAT genes in 14 tissues. The transcriptome data of 14 tissues at three stages was used to reconstruct expression patterns of wheat AAT genes. The tissues at the seeding stage included radicle (SRA), coleoptile (SC), stem axis (SSA), first leaf blade (SFL), roots (SR) and shoot apical meristem (SSAM); the tissues at the flag leaf stage included flag leaf blade (FFL), shoot axis (FSA), roots
(FR); and the tissues at the milk grain stage included flag leaf blade (MFL), peduncle (MP), awns (MA), glumes (MGL), and grain (MG). Samples are listed at the bottom of each lane, and the color scale is shown at the right.

**Fig. 9. Three trends of expression patterns of duplicated AAT gene pairs in wheat.** The X-axis indicates 14 tissues at three stages, and the Y-axis represents the TPM value. The full names for X-axis tissue abbreviations are shown in Fig. 8.

**Fig. 10. The expression levels of 12 selected TaAAT genes in leaves and roots under three abiotic stresses.** (a) The relative expression levels of TaAATs under drought (D) and heat stresses (H) for 1 h and 6 h in leaves at the three-leaf stage. (b) The relative expression levels of TaAATs under salt stress (S) (NaCl, 200 mM) for 6 h, 12 h, 24 h and 48 h in roots at the three-leaf stage. Bars represent the mean values of three replicates ± standard deviation (SD). All of the expression levels of the TaAAT genes were normalized to the expression level of TaActin.
Supplementary information

Table S1. The general information and sequence characterization of 296 TaAAT genes.

Table S2. The primers used for quantitative real time polymerase chain reaction (qRT-PCR).

Figure S1. The conserved motifs of all AAT genes in wheat.
The English in this document was edited for proper English language, grammar, punctuation, spelling, and overall style by the highly qualified native English speaking editors at American Journal Experts. The Certificate Verification Key is 8F62-3901-5164-55BF-6277.