

# SCIENTIFIC REPORTS



OPEN

## Massive expansion and differential evolution of small heat shock proteins with wheat (*Triticum aestivum* L.) polyploidization

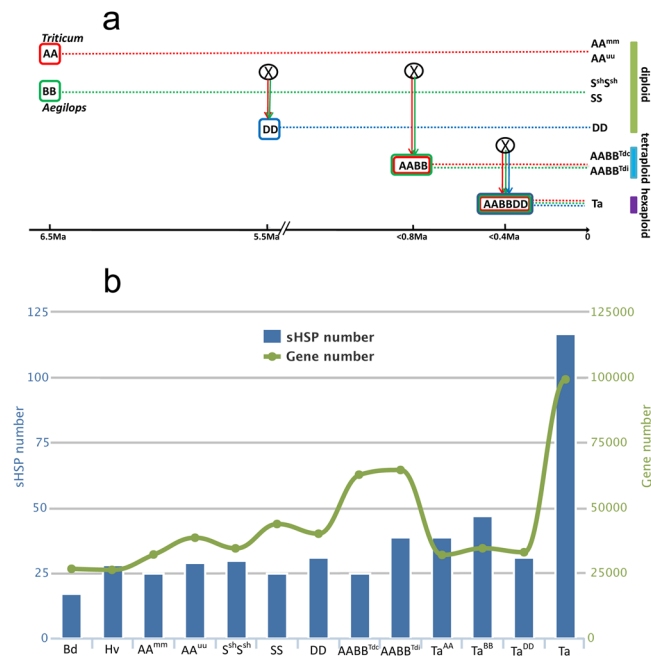
Xiaoming Wang<sup>1</sup>, Ruochen Wang<sup>1</sup>, Chuang Ma<sup>1,2</sup>, Xue Shi<sup>1</sup>, Zhenshan Liu<sup>1</sup>, Zhonghua Wang<sup>1</sup>, Qixin Sun<sup>1,3</sup>, Jun Cao<sup>4</sup> & Shengbao Xu<sup>1</sup>

Wheat (*Triticum aestivum*), one of the world's most important crops, is facing unprecedented challenges due to global warming. To evaluate the gene resources for heat adaptation in hexaploid wheat, small heat shock proteins (sHSPs), the key plant heat protection genes, were comprehensively analysed in wheat and related species. We found that the sHSPs of hexaploid wheat were massively expanded in A and B subgenomes with intrachromosomal duplications during polyploidization. These expanded sHSPs were under similar purifying selection and kept the expressional patterns with the original copies. Generally, a strong purifying selection acted on the  $\alpha$ -crystallin domain (ACD) and theoretically constrain conserved function. Meanwhile, weaker purifying selection and strong positive selection acted on the N-terminal region, which conferred sHSP flexibility, allowing adjustments to a wider range of substrates in response to genomic and environmental changes. Notably, in CI, CV, ER, MI and MII subfamilies, gene duplications, expression variations and functional divergence occurred before wheat polyploidization. Our results indicate the massive expansion of active sHSPs in hexaploid wheat may also provide more raw materials for evolving functional novelties and generating genetic diversity to face future global climate changes, and highlight the expansion of stress response genes with wheat polyploidization.

Diverse heat shock proteins (HSPs), one key component of cellular protein quality control network, are very important for plant heat stress adaptation<sup>1-3</sup>. Small heat shock proteins (sHSPs), one type of HSPs, are typically induced by heat, and function in protein folding and preventing or reversing substrate aggregation<sup>3,4</sup>. sHSPs contain a conserved  $\alpha$ -crystallin domain (ACD), preceded by a highly divergent N-terminal region and followed by a C-terminal extension. The ACD comprises seven or eight anti-parallel  $\beta$ -strands which form a  $\beta$ -sandwich. In their native state, sHSPs form oligomers (200–800 kDa) that operate as a functional unit<sup>5,6</sup>. The ACD and C-terminal extension are the major regions responsible for oligomer formation, while the flexible N-terminal region is responsible for substrate recognition and oligomer stability<sup>3,4,7-11</sup>. Extensive studies have demonstrated that manipulating the sHSPs of plants remarkably altered their thermotolerance<sup>12-16</sup>, highlighting their role in heat adaptation. The composition and expression of sHSPs has recently been studied in several plants, including *Arabidopsis thaliana*, rice, algae, *Populus trichocarpa*, pepper, sugarcane and soybean, and the identified sHSPs have been categorized into sixteen subfamilies, including eleven cytoplasmic/nuclear localized (CI-CXI) and five organelle localized subfamilies<sup>17-24</sup>. However, the evolutionary mechanisms of sHSPs in plants, especially in plant polyploidization process, remain largely unknown.

Productivity of bread wheat (*Triticum aestivum* L.) cropping systems is at risk due to increasing temperatures as a result of global warming<sup>25-27</sup>. Bread wheat, also known as allohexaploid wheat (AABBDD), originated from two hybridization events among the genera *Triticum* and *Aegilops*<sup>28,29</sup>. The first hybridization, between *Triticum urartu* (AA) and a close relative *Aegilops speltoides* (BB), occurred <0.8 million years ago (Ma) and gave rise to

<sup>1</sup>State Key Laboratory of Crop Stress Biology for Arid Areas, College of Agronomy, Northwest A&F University, Yangling, Shaanxi, 712100, China. <sup>2</sup>College of Life Sciences, Northwest A&F University, Shaanxi, 712100, China. <sup>3</sup>Department of Plant Genetics & Breeding, China Agricultural University, Yuanmingyuan Xi Road No. 2, Haidian District, Beijing, 100193, China. <sup>4</sup>Innovation Experimental College, Northwest A&F University, Shaanxi, 712100, China. Correspondence and requests for materials should be addressed to S.X. (email: xushb@nwsuaf.edu.cn)



**Figure 1.** Number of sHSPs in wheat and related species. **(a)** Evolutionary history of bread wheat. The symbol ‘X’ in circle mark hybridisation event. This figure is modified from [Marcussen, T. *et al.* Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345 (2014)]<sup>30</sup>. Reprinted with permission from AAAS. **(b)** Numbers of sHSPs in different species. Left Y axis represents the identified sHSP numbers and right Y axis represents the gene numbers contained in the genome sequences analysed. Abbreviations: Ma, million years ago; Bd, *Brachypodium distachyon*; Hv, *Hordeum vulgare*; AA<sup>mm</sup>, *Triticum monococcum*; AA<sup>uu</sup>, *Triticum urartu*; S<sup>sh</sup>S<sup>sh</sup>, *Aegilops sharonensis*; SS, *Aegilops speltoides*; DD, *Aegilops tauschii*; AABB<sup>tdc</sup>, *Triticum durum*, cv. Strongfield; AABB<sup>tdi</sup>, *Triticum durum*, cv. Cappelli; Ta<sup>AA</sup>, subgenome A of bread wheat; Ta<sup>BB</sup>, subgenome B of bread wheat; Ta<sup>DD</sup>, subgenome D of bread wheat; Ta, *Triticum aestivum*.

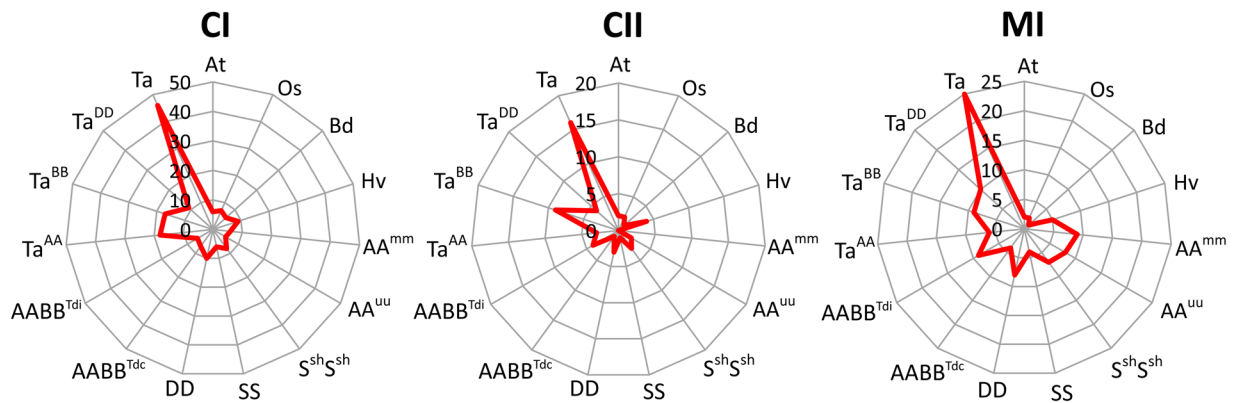
the allotetraploid emmer wheat (*Triticum turgidum*; AABB). The second hybridization, between emmer wheat and *Aegilops tauschii* (DD), which arose from the hybridization of the AA and BB lineages ~5.5 Ma, occurred <0.4 Ma and gave rise to the allohexaploid wheat<sup>30</sup>. Owing to the high degree of autonomy in the subgenomes and the shorter evolutionary history<sup>31</sup>, bread wheat and its related progenitors provide a good system for studying the polyploidization process<sup>30</sup>.

Recent studies have suggested that all flowering plants have undergone at least two ancient polyploidy events during their evolutionary history<sup>32</sup>. The autopolyploidy which resulted from whole genome duplication or unreduced gametes fusion within a single species, and the allopolyploidy which resulted from interspecific hybridisation, generate extra gene copies<sup>33,34</sup>. These duplicated genes then underwent subfunctionalization, neofunctionalization, specialization, pseudogenization or concerted evolution, leading to greater genetic diversity<sup>34–36</sup>, which is the key factor in adaptation to new habitats and distribution to larger geographical areas than those covered by their progenitors<sup>28,34</sup>. The genome size of hexaploid wheat approximates to the sum of three diploid progenitors with 2–10% DNA loss with polyploidization, suggesting that most genes are redundant compared with their progenitors and some genes may be lost<sup>28</sup>. As the key heat protection genes, twenty-seven candidate sHSPs have been identified in hexaploid wheat, with all found to localise to the mitochondria (26) or nucleus (1)<sup>37</sup>. This differed from the literature on the cellular localization of other identified sHSPs, which has shown that most sHSPs are actually located in the cytoplasm<sup>17,20,22</sup>. Further elucidation and clarification of sHSP candidates in wheat is therefore required.

With the accumulation of genome data on hexaploid wheat and its relatives, systematic analyses of sHSP evolution and function in Triticeae have become practicable. Here, we demonstrate the evolutionary pattern and mechanism of sHSPs in Triticeae and provide evidence for massive intrachromosomal gene duplications, selection pressure variations and the expression pattern diversity during wheat polyploidization. Our results provide fundamental evidence that will be valuable for further wheat sHSP applications as well as for its evolutionary pattern variations with polyploidization.

## Results

**Massive expansion of sHSPs during wheat polyploidization.** To characterize the copy number variation of sHSPs during wheat polyploidization, we performed comparative analyses on hexaploid wheat and its relatives (diploid and tetraploid). The evolutionary relationships among Triticeae were constructed (Fig. 1a), with reference to the results of Marcussen *et al.*<sup>30</sup>. Diploid progenitors were represented by *T. urartu* (AA<sup>uu</sup>) and *Triticum monococcum* (AA<sup>mm</sup>) as the wheat subgenome A donors, *Ae. tauschii* (DD) as the wheat subgenome D donor, and *Ae. sharonensis* (S<sup>sh</sup>S<sup>sh</sup>) and *Ae. speltoides* (SS) as the wheat subgenome B donors. Tetraploid



**Figure 2.** Copy number variation of CI, CII and MI subfamily during wheat polyploidization. Abbreviations, At, *Arabidopsis thaliana*; Os, *Oryza sativa*; Bd, *Brachypodium distachyon*; Hv, *Hordeum vulgare*; AA<sup>mm</sup>, *Triticum monococcum*; AA<sup>uu</sup>, *Triticum urartu*; S<sup>sh</sup>S<sup>sh</sup>, *Aegilops sharonensis*; SS, *Aegilops speltoides*; DD, *Aegilops tauschii*; AABB<sup>Tdc</sup>, *Triticum durum*, cv. Strongfield; AABB<sup>Tdi</sup>, *Triticum durum*, cv. Cappelli; Ta<sup>AA</sup>, subgenome A of bread wheat; Ta<sup>BB</sup>, subgenome B of bread wheat; Ta<sup>DD</sup>, subgenome D of bread wheat; Ta, *Triticum aestivum*.

progenitors were represented by *T. turgidum* (Cappelli, AABB<sup>Tdc</sup>) and *T. turgidum* (Strongfield, AABB<sup>Tdi</sup>) as the reference AABB genomes<sup>29–31, 38–41</sup>.

With reference to the sHSP identifications in *A. thaliana* and rice<sup>17, 20, 22</sup>, 404 sHSP candidates were identified in Triticeae, including 321 sHSP genes and 83 *Acd* genes (which shared homology with the ACD domain of sHSPs but were divergent from sHSPs), respectively (Supplementary Fig. S1). About 25 to 31 sHSP genes were identified in seven diploid and tetraploid relatives (Fig. 1b, Supplementary Table S1), suggesting that this stable amount of sHSPs may be sufficient to maintain the heat adaptation capability of Triticeae. Surprisingly, 117 sHSPs were identified in bread wheat, many more than in its relatives, even remarkably more than the total (56–70) of a tetraploid (AABB) and the DD progenitors, or the sum (81–90 in total from AA, BB and DD) of its three diploid progenitors (Fig. 1b, Supplementary Table S1). In a previous study, the copy number variation of wheat genes was analysed, which showed that the gene numbers of all six investigated gene families in hexaploid wheat were considerably less than the sum of the three diploid progenitors<sup>31</sup>. These results indicate that the sHSPs, as environmental adaptation genes, were unusual and remarkably expanded in hexaploid wheat with polyploidization.

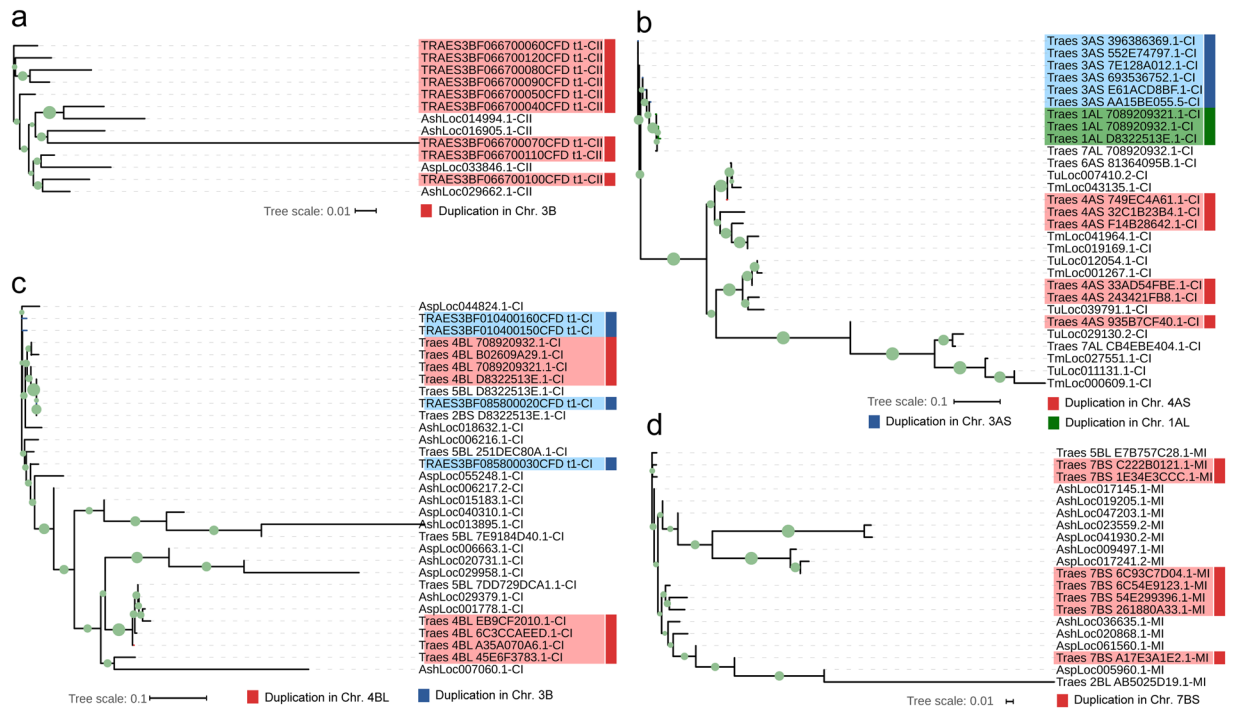
Tetraploid wheat (AABB<sup>Tdc</sup> and AABB<sup>Tdi</sup>) originated recently (around 0.8 Ma)<sup>30</sup>. Our result shows that the genome size of current tetraploid is close to the total of its two diploid progenitors, along with 13.1% (AA<sup>uu</sup> and S<sup>sh</sup>S<sup>sh</sup> to AABB<sup>Tdi</sup>)–31.2% (AA<sup>uu</sup> and SS to AABB<sup>Tdc</sup>) gene loss (Fig. 1b). However, the number of sHSPs in tetraploid wheat was similar to that in its diploid progenitors, rather than reflecting the sum of the two diploid progenitors (Fig. 1b). In hexaploid wheat, the genome size approximates to the sum of the three diploid progenitors, along with 2–10% DNA loss<sup>28</sup>. The number of sHSPs in hexaploid wheat, however, was much higher than the sum of its three diploid progenitors (Fig. 1b). Further genomic comparison revealed remarkable sHSPs expansion in subgenomes A and B, but not in subgenome D (Fig. 1b, Supplementary Table S1), in consistent with that the genes in subgenome D keep relatively stable with polyploidization<sup>42–44</sup>.

**The sHSPs of hexaploid wheat expanded in the CI and CII subfamilies with intrachromosomal duplications.** To further understand sHSP expansion during wheat polyploidization, the sHSPs of Triticeae were classified into 13 known subfamilies (Supplementary Fig. S1). No sHSPs belonging to the CIV, CVII and CXI subfamilies were detected among Triticeae. However, interestingly, our analysis identified a new subfamily specific to Triticeae that was designated as CXII onwards.

The subfamilies CI, CII contained the most members among the 14 identified sHSP subfamilies in Triticeae. The sHSPs belonging to these two subfamilies also exclusively showed remarkable expansion in hexaploid wheat compared with the total number of sHSPs among the three diploid progenitors, or among the tetraploid and DD progenitors (Fig. 2, Supplementary Table S1). Gene expansion in subgenome A was mainly associated with the CI subfamily, while gene expansion in subgenome B was mainly associated with the CII subfamilies.

Interestingly, these expanded sHSPs were enriched in specific chromosome fragments. For example, remarkable CII sHSP expansion occurred in subgenome B (Fig. 2, Supplementary Table S1), and all nine CII sHSPs in subgenome B were located in a 1 Mb region of chromosome 3B (Supplementary Fig. S2), suggesting the expanded sHSPs may be the result of intrachromosomal duplications. Further sequence comparative analysis demonstrated that these nine genes expanded from 1–3 original sHSPs during polyploidization (Fig. 3a). Similarly, sHSP expansion in the CI subfamily particularly in chromosomes 4AS, 3AS, 1AL (Fig. 3b), 4BL and 3B (Fig. 3c), and the sHSPs of the MI subfamily mainly located in chromosome 7BS (Fig. 3d). These results suggest that sHSP expansion in hexaploid wheat originated from intrachromosomal duplications, which is consistent with the segmental duplications of sHSPs observed in *A. thaliana*, rice and soybean<sup>22, 23, 45</sup>, and exhibit a notably higher intrachromosomal duplication ratio during wheat polyploidization<sup>31, 41</sup>.

However, not all sHSP subfamilies displayed expansion in hexaploid wheat. Subfamilies CIII, CX and CXII exhibited decreased numbers of sHSPs compared with the diploid or tetraploid progenitors (Supplementary



**Figure 3.** Gene expansions of subfamilies CI, CII and MI in subgenomes A and B. **(a)** Duplications of subfamily CII in subgenome B. **(b)** Duplications of subfamily CI in subgenome A. **(c)** Duplications of subfamily CI in subgenome B. **(d)** Duplications of subfamily MI in subgenome B. The bootstrap values are plotted as circles at the nodes with the circle size proportional to the bootstrap value. Abbreviations: Asp, *Aegilops speltoides*; Ash, *Aegilops sharonensis*; Tm, *Triticum monococcum*; Tu, *Triticum urartu*; Traes, *Triticum aestivum*.

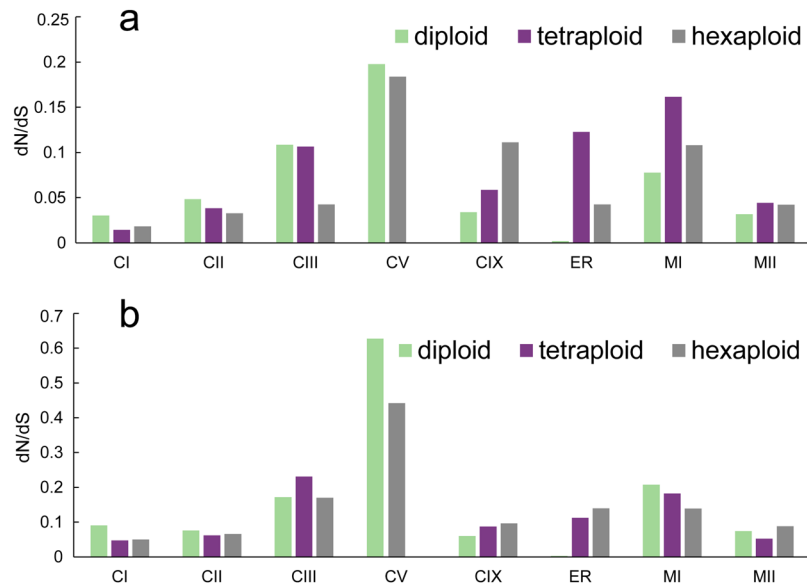
(Fig. S3). This result suggested that sHSP loss should not be overlooked as it occurs simultaneously with gene expansion during wheat polyploidization.

**Severe purifying selection acts on sHSPs during wheat polyploidization.** Previous studies have demonstrated that different subfamilies of sHSPs possess distinct evolution histories in plants<sup>4</sup>. To unveil the underlying evolutionary mechanism of each sHSP subfamily in Triticeae and the natural selection acted on duplicated sHSPs in hexaploid wheat, we estimated the ratio ( $\omega$ ) of nonsynonymous (dN) to synonymous (dS) substitution rates ( $\omega = dN/dS$ ) with codon-based models provided in the PAML programs<sup>46</sup>. The  $\omega$  values estimated by the one-ratio model were 0.05 and 0.1 for the ACD domain and the full length proteins, respectively, indicating that strong purifying selection acted on this protein family to remove nonsynonymous mutations and to maintain their important biological functions (Fig. 4). In general, different sHSP subfamilies were under different types of purifying selection, and the  $\omega$  of each subfamily displayed similar purifying selection between the ACD domain and the full sHSP protein sequence, but weaker selection on the full sHSP sequence. This result is consistent with the fact that the conserved ACD domain is the core region for maintaining sHSP function.

The subfamilies CI, CII and P are the most ancient sHSPs that arose around 400 Ma, while other sHSPs evolved from these three subfamilies much more recently<sup>47, 48</sup>. Among the diploid progenitors, the ACD domain of the CI, CII and P subfamilies showed lower  $\omega$  values (Fig. 4). In addition, the CIX and MII subfamilies also showed strong purifying selection compared with other subfamilies, such as the CIII, CV and MI subfamilies, which were under relatively weaker purifying selection. This finding was consistent with previous reports that CV sHSPs showed distinct expression patterns in *A. thaliana* and rice<sup>20, 22</sup>, indicating that they had experienced a function shift with weaker selection pressures.

In tetraploid relatives and hexaploid wheat, similar selection patterns were observed, indicating that these patterns were conserved across Triticeae relatives. For CI and CII subfamilies which experienced gene duplications, similar purifying selections were observed between hexaploid wheat and its progenitors, indicating that the gene duplications did not alter the selection pressures acted on these two subfamilies and the duplicated genes also under strong purifying selections. However, a stronger purifying selection was observed for the subfamily CIII in hexaploid wheat, while obvious weaker purifying selection acted on subfamilies of CIX and P (Fig. 4), suggesting that subfamilies CIII, CIX and P may experience altered selection pressure during the polyploidization process.

Taken together, these findings indicate that specific purifying selection is acting on each sHSP subfamily and is conserved across Triticeae relatives, which may lead different sHSP subfamilies in distinct evolutionary directions. Although massive gene expansions occurred, the sHSPs in hexaploid wheat have evolved in this manner, with the exception of minor differences in the CIII, CIX and P subfamilies.



**Figure 4.** Selection pressure variation during wheat polyploidization. **(a)** The selection pressures acting on the ACD domain region. **(b)** The selection pressures acting on the full sHSPs. The X axis represents the different sHSP subfamilies and Y axis represents the dN/dS value. The dN/dS values were estimated based on the total sHSPs of diploid, tetraploid and hexaploid genomes, respectively. If the sHSP number was lower than three, the dN/dS value was not calculated.

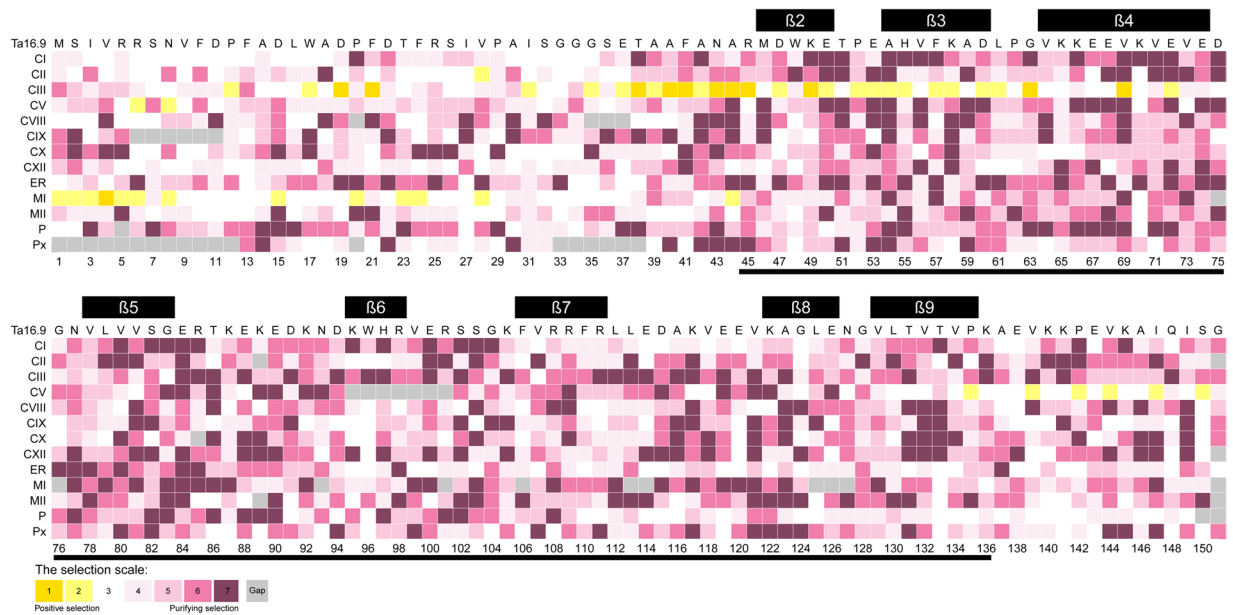
**The sHSPs of hexaploid wheat were duplicated in divergent sHSP groups.** Gene duplications, arising from single gene duplication or whole genome duplication events, result in enlarged and potentially more complex gene families. In this study, the sHSPs of the CI, CV, CIX, ER, MI and MII subfamilies could be divided into two groups evidently in the polygenetic tree, and these two groups were likely paralogs originating from gene duplications (Supplementary Fig. S4). These duplicated sHSPs showed obvious species preferences; for example, the groups CIb and ERb were only present in monocots, whereas the groups CIXb, MIb and MIIb were unique to Triticeae. These results indicated that the original duplications, which led to the different groups of sHSPs, were ancient events that mostly related to speciation, rather than to the wheat polyploidization.

To understand the evolutionary fates of duplicated genes, the codon-based models contained in PAML<sup>46</sup> were employed (Supplementary Table S2). Analysis using the one-ratio model (model 0) and the two-ratio branch model (model 2) showed that the two groups of CIX, ER and MI subfamilies had different  $\omega$  values, indicating asymmetrical evolution between the duplicated groups. The site-specific discrete model (model 3) and the clade model (model D) were applied to detect divergent selective pressure. The results demonstrated that the selective pressures that acted on the two groups of the CI, CV, ER, MI and MII subfamilies were distinctly different (Supplementary Table S2), thus suggesting that functional divergence occurred between paralogs. Furthermore, many type I sites (conserved in one duplicated cluster but highly variable in another duplicated cluster) and type II sites (highly conserved in two duplicated clusters but variable between them) responsible for functional divergence were identified (Supplementary Fig. S5). In general, the number of type II sites identified was greater than the number of type I sites (only type II sites were identified in the CV, ER and MII subfamilies), and these sites were mainly located in the ACD regions. This finding further supported the theory that the two groups of these subfamilies had experienced substantial functional divergence.

To test the effect of positive selection on functional divergence, the branch-site model A (model A) and the null model A were applied (Supplementary Table S2). Different positive selection sites were detected between the leading branches of CIa/CIb and ERa/ERb. However, two duplicated groups of CI and ER, displayed the same positive selection sites, suggesting that positive selection occurred before gene duplication and was not therefore the reason for functional divergence.

In conclusion, the duplicated clades of the CI, CV, ER, MI and MII subfamilies were under distinct selection pressure acting on the ACD domain, which led to functional divergence. However, such gene duplications and functional divergences were ancient events, occurring long before hexaploid wheat polyploidization. Furthermore, the expansion of sHSPs in hexaploid wheat occurred in both of the diverged groups, suggesting that the expansion of sHSPs was not related to the functional divergence of sHSPs, but simply to an increase in the numbers in the diverged sHSP group.

**Weaker purifying selection and evident positive selection in the N-terminal region conferred sHSPs with more flexibility.** To evaluate the selection pressures that acted on each region of the sHSPs in Triticeae, we aligned the sHSP sequences of each subfamily to Ta16.9, the model used for sHSP studies<sup>6</sup>, to detect the selection pressures that acted on each amino acid residue (Fig. 5). The results showed that no conserved residue was under strong purifying selection across all subfamilies, even in the  $\beta$ -strands, with the most conserved region being in the ACD<sup>3</sup>. Generally, the residues in the ACD and the C-terminal extension regions were under



**Figure 5.** Selection pressures acting on each residue of the sHSPs. The protein sequences of each subfamily were aligned to the sHSP 16.9 of *Triticum aestivum* (Ta16.9), for which the crystal structure has been reported<sup>6</sup>, and selection pressures acting on each residue were estimated in each subfamily. The black line at the bottom marks the ACD domain. The protein sequence of Ta16.9 and its secondary structure ( $\beta$ -strands) is shown at the top. The selection scale is shown at the bottom.

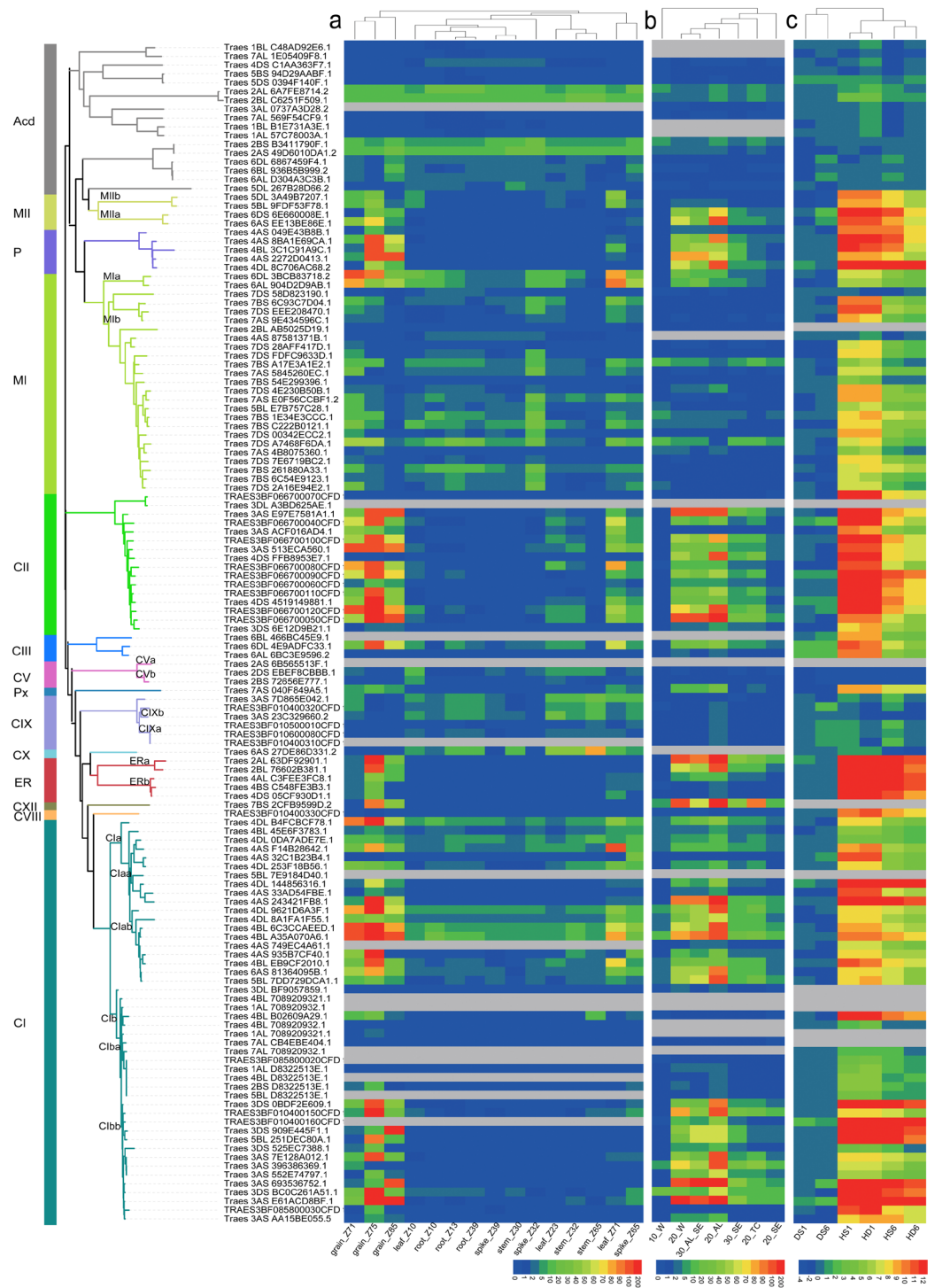
stronger purifying selection than those in the N-terminal region. By contrast, the positive selection sites identified were mainly located in the N-terminal region of the CIII (eight of 11 residues located in the N-terminal) and MI (one residue located in the N-terminal) subfamilies. The N-terminal region plays a major role in substrate recognition and binding<sup>3,4,7–11</sup>. Thus, the weaker purifying selection and the evident positive selection in the N-terminal region may confer greater flexibility in sHSPs, allowing the binding of more substrates.

**The sHSPs of hexaploid wheat are actively transcribed.** The expression profiles, as proxies for function, provide a valuable resource to study the evolutionary trajectories of duplicated genes<sup>34</sup>. To test the evolutionary patterns of sHSPs estimated by the above sequence analysis in this study, an integration transcriptional analysis was performed on published RNA-seq datasets from 28 hexaploid wheat samples, covering five tissues, three developmental stages and two types of abiotic stress (Fig. 6). In total, 95.7% (112 out of 117) of the identified sHSPs from bread wheat had an FPKM value higher than zero for at least one condition (6–15 samples for each condition), and were regarded as transcribed sHSPs, a much higher number of transcribed genes than in the wheat genome (71.4% on average)<sup>41</sup>. This result indicates that sHSPs (including expanded sHSPs) of hexaploid wheat are highly active and massive pseudogenization did not occur after polyploidization (Fig. 6), potentially contributing to recent wheat genome evolution and, therefore, to the adaptive evolution of bread wheat.

The expression of these active sHSPs showed obvious spatiotemporal preferences (Fig. 6a,b). In general, the sHSPs were highly expressed at late developmental stages and in the corresponding tissues (spikes, filling grains and leaves after anthesis), which was consistent with the fact that frequent heat stress usually occurred after anthesis in most wheat growth areas<sup>49</sup>. The higher expression levels of sHSPs may enable rapid responses to heat stress or pre-heat stress. By contrast, some subfamilies displayed a constitutive expression pattern in all tissues and at all developmental stages. These subfamilies, which included MI, CIX and CV, showed low-level selection pressure and were less sensitive to heat induction. These results suggest that lower selection pressure may be related to the functional constraint shifts of constitutively expressed sHSPs.

As mentioned above, the duplicated groups of the CI, CV, ER, MI and MII subfamilies had undergone functional divergence. Consistent with this, the diverged duplicated groups of CI, ER, MI and MII showed distinct expression patterns (Fig. 6a,b, Supplementary Table S3). Only one group from each subfamily (ERa, MIa and MIIa) was found to be abundantly expressed in grain at 20 and 30 days post-anthesis. Similarly, in the CI subfamily, each of the duplicated groups could be further divided into two subgroups (CIaa, CIab and CIba, CIbb), and each subgroup showed distinct expression patterns. These results provide expression-level evidence of functional divergence between these duplicated groups. It should be noted that the expansion of hexaploid sHSPs occurs independently to the transcription patterns of these genes, i.e. less highly transcribed sHSPs are duplicated as frequently as more highly transcribed sHSPs.

Heat induction is a characteristic feature of sHSPs, and the RNA-seq data showed that 81.2% of sHSPs were markedly induced by heat (fold change  $\geq 2$  and false discovery rate (FDR) adjusted  $p < 0.01$ ), a similar proportion as seen with *A. thaliana* and rice sHSPs (87–90%)<sup>20,22</sup>. This finding indicates that expanded sHSPs in the current wheat genome maintain their conserved function in heat protection (Fig. 6c). Consistent with the different



**Figure 6.** Phylogenetic tree of wheat *sHSP* genes and their expression profiles. The phylogenetic tree of wheat *sHSP* and *Acd* (that shared homology with the *ACD* domain of *sHSP*s but diverged from *sHSP*s) genes is shown on the left side of the panel. The clades of each subfamily are depicted by specific colours, and the names of subfamilies are marked to the left. The symbols of duplicated groups derived from Supplementary Fig. S4 were marked at the leading clades of each group. **(a)** The expression profiles were estimated using RNA-seq data from five organs at three developmental stages. Transcript levels are depicted using a colour scale indicating FPKM (fragments per kilobase of exon model per million mapped reads) values. The symbols representing developmental stages were as previously reported<sup>31</sup>. **(b)** The expression profiles were estimated using RNA-seq data from different grain cell types at three developmental stages. Transcript levels are depicted using a colour scale indicating the FPKM values. The symbols representing cell types and developmental stages were as previously reported<sup>29</sup>. **(c)** The expression profiles under heat (HS), drought (DS) and the combined stress of heat and drought (HS) after 1 and 6 h treatments<sup>65</sup>. Transcript levels are depicted using a colour scale indicating  $\log_{10}$  values. The grey lines in panels a, b and c indicate that the *sHSP* genes were not detected in the corresponding RNA-seq dataset.

selection pressures and functional divergence, the expression patterns of different subfamilies and duplicated groups varied considerably under heat stress. The expression of CI, CII, P, ER, MII and Px, which were subject to stronger purifying selection, were up-regulated sharply, whereas the expression of CV, CIX and MI, which were subject to lower levels of purifying selection, were either not induced or only weakly induced by heat stress (Fig. 6c). The expression levels of duplicated groups of CI and MII also varied considerably. Among the expansion subfamilies of CI and CII, only CII (mainly possessing expansion on chromosome 3B) and parts of CI were strongly induced by heat stress (Fig. 6c), and similar transcriptional profiles were observed in grains. In fact, heat-induced sHSPs are also highly expressed in grains (Fig. 6), indicating that these sHSPs operate as a core heat response system in hexaploid wheat. In response to common abiotic stress, the majority of sHSPs showed strong heat induction, while a few were induced by drought stress, although similar effects resulted from both. To confirm the heat induction characters in transcriptional analysis, further qRT-PCR analysis showed that sHSPs in seedling leaves were rapidly induced by heat stress, and showed similar time-course heat response profiles (Supplementary Fig. S6 and Table S4), highlighting that sHSPs of bread wheat played their key roles in wheat adaptation to heat environments.

The transcriptional profiles and tissue preference of sHSPs displayed strikingly consistency with the evolutionary analysis based on sequences variations, further supporting the purifying selection and differential evolution. In addition, the duplicated sHSPs of CI and CII subfamilies kept similar transcripton patterns with original sHSPs ( $t$ -test  $p > 0.05$ ,  $p = 0.1354$  for CI and  $p = 0.1348$  for CII, respectively), consistent with the unaltered purifying selections acted on them.

## Discussion

Here, we performed systematic sHSP evolution and expression analyses in Triticeae and revealed remarkable sHSP expansion along with wheat polyploidization. We also characterised the current sHSP status in bread wheat and hypothesised about future evolutionary patterns.

A previous study suggested that the genome size of hexaploid wheat was approximately the sum of its three diploid progenitors with 2–10% DNA loss<sup>28</sup>, indicating that, in general, each gene may be amplified three-fold in hexaploid wheat compared with their diploid progenitor. However, a recent study showed that the members of six gene families was considerably less than the sum of the three diploid progenitors<sup>31</sup>, suggesting that significant gene loss had occurred in these gene families with polyploidization, and some yet to be investigated gene families must have expanded to nearly triple the genome size of the diploid progenitors. In the current study, sHSPs were found to be dramatically expanded in hexaploid wheat, to an extent greater than merely the sum of its progenitors, which was consistent with the fact that stress response genes were over-represented with polyploidization in chromosome 3B<sup>41</sup>. These results suggested that the considerable expansion of a few gene families is also a common characteristic of wheat polyploidization, as well as the significant gene loss in specific gene families.

To date, only chromosome 3B has been sequenced completely<sup>41</sup>. The sHSP subfamily CII is dramatically expanded in this chromosome, and the sequence of its numbers share high similarity (Fig. 3a). Considering the short polyploidy history of tetraploid and hexaploid wheat<sup>30</sup> and the corresponding high sequence similarity of expanded sHSPs, some expanded sHSPs with polyploidization may be undetectable by whole-genome or chromosome-based shotgun sequencing approaches, as previously reported<sup>31</sup>. Thus, the amounts of sHSPs in tetraploid and hexaploid wheat are likely to have been underestimated. Similarly, the degree of purifying selection acting on sHSPs may have been overestimated because of the limited numbers of mutations accumulated during the short evolutionary history of hexaploid wheat.

In this study, sHSPs duplications are specific occurred in subgenome A and B, but not D. The subgenome D, as a new member of hexaploid genome<sup>30</sup>, displayed quiet and stable status during the polyploidization. In addition, the sHSPs in subgenome D keep the conserved sequence and classification with their diploid progenitor, and show similar purifying selection pressure with its diploid progenitors. With the improvement of wheat genomics, several conclusions about subgenomic biases in wheat development and stress responses were reported, such as the cell type and stage dependent subgenome dominance in wheat grain development<sup>29</sup>, the bias towards D subgenome in response to *Fusarium* head blight<sup>50</sup>, towards B and D subgenomes in response to *Fusarium pseudograminearum*<sup>51</sup>.

Diploidization makes the loss of duplicated genes and returns the most fractions of a polyploid genome back to the singleton state<sup>32</sup>. Our result shows that the genome size of current tetraploid is close to the total of its two diploid progenitors, along with 13.1% (AA<sup>uu</sup> and S<sup>sh</sup>S<sup>sh</sup> to AABB<sup>Tdi</sup>)–31.2% (AA<sup>uu</sup> and SS to AABB<sup>Tdc</sup>) gene loss (Fig. 1b), suggesting that the tetraploid genome has not experienced or is still undergoing the diploidization process. However, the number of sHSPs in tetraploid wheat was similar to that in its diploid progenitors, rather than reflecting the sum of the two diploid progenitors (Fig. 1b), indicating that a remarkable number of sHSPs had been lost in tetraploid wheat, likely has experienced diploidization process. In addition, it is reported that the allotetraploidization in wheat rapidly induced extremely biased homeolog expression and this trend is further enhanced in the subsequential domestication and evolution of polyploid wheats, providing an indirect evidence for the diploidization of tetraploid wheat<sup>34,52</sup>, and supporting the diploidisation may be responsible for unusual less sHSPs in tetraploid wheat. However, more genomic complexity and incomplete genome sequences of tetraploid wheat still be the potential reasons for the smaller number of sHSPs in tetraploid.

Although both inter- and intrachromosomal duplication rates are apparently higher in wheat than in other grass species<sup>31,41</sup>, expanded sHSPs in hexaploid wheat mainly resulted from intrachromosomal duplications in this study, highlighting the role of segmental or tandem duplications in gene expansion with wheat polyploidization. Interestingly, genes involved in the stress response were also identified in the typically overrepresented gene category in chromosome 3B<sup>41</sup>, indicating that sHSP expansion may not occurred at random, but instead occurred specifically with wheat polyploidization. It will be intriguing to investigate why stress response genes particularly



expanded, as well as the mechanism by which organisms expand specific genes by intrachromosomal duplications during polyploidization.

Selection pressure is the key factor driving speciation and genetic evolution. In this study, we revealed that purifying selection and positive selection acted via a conserved pattern on sHSPs across members of Triticeae, indicating that the sHSPs of hexaploid wheat, despite high redundancy, adhere to the common evolutionary rules of Triticeae sHSPs. The weaker levels of purifying selection and positive selection on the N-terminal region of sHSPs may allow for greater functional plasticity, potentially allowing these proteins to adjust to their substrates, preventing or reversing the unnatural substrate aggregations usually caused by heat stress<sup>3,4,7–11</sup>. For hexaploid wheat, this may be a critical step for sHSP evolution in polyploidization, because sHSPs may not only need to recognize the original substrates from the same subgenome, but also need to recognize the homolog or heterolog substrates from other subgenomes, which may be critical for future diploidization.

Owing to the dosage and functional redundancy of sHSPs in wheat, greater complexity and diversity at the evolutionary and functional level were observed than the sHSPs in related species (Figs 1, 2 and 6). Theoretically, this redundancy could provide a larger reservoir for genetic variation, facilitating the evolution of new sHSPs or novel functions, a more efficient response to heat stress and a stronger adaptability to wide range of climatic conditions, which was assumed to be a key factor in the success of wheat as a global food crop<sup>31</sup>. Paradoxically, bread wheat is actually heat vulnerable<sup>25,53–55</sup>. This fact suggests that hexaploid wheat thermotolerance can't be enhanced by simply increasing sHSPs amount. However, the expansion of sHSPs prefer provide massive genome resources to compensate the weak thermotolerance. This hypothesis may provide a logic that why stress response genes prefer be expanded with polyploidization<sup>41</sup>. On the other hand, due to the early polyploidization status of bread wheat<sup>31</sup>, the interaction between wheat genome and environmental changes and the acquirement of heat adaption may be ongoing.

Most of the expanded sHSPs in hexaploid wheat are likely to be lost in the diploidization process when it adapts to its habitat and corresponding environment<sup>28</sup>, as demonstrated by our findings in the current diploid and tetraploid relatives. Considering that global warming is an unprecedented event in the evolutionary history of Triticeae, the future evolution of sHSPs is difficult to predict. However, it is clear that hexaploid wheat has evolved a massively expanded number of active sHSPs with polyploidization, and is ready to action in new heat environments.

In summary, our study presents the evidence of the considerable expansion of sHSPs by intrachromosomal duplications in hexaploid wheat. Of these abundant sHSPs, the ACD domain is under strong purifying selection to maintain conserved function, while the highly variable N-terminal region confers greater plasticity for substrates binding to response to genomic and environmental changes. Most sHSPs maintain active transcription and the ability for heat induction. We speculate that the expansion of active sHSPs with wheat polyploidization confers extra gene resource for the future adaption evolution of bread wheat, like the sHSPs duplications in CI, CV, ER, MI and MII subfamilies, subsequently functional divergence, and corresponding expression variations occurred before wheat polyploidization. Wheat researchers and breeders may therefore make full use of sHSP resources to improve wheat heat tolerance in the face of global warming. Our findings also provide insight into the intrachromosomal duplications, gene expansions and environmental adaptation processes that occurred along with plant polyploidization.

## Methods

**Data collection and identification of Triticeae sHSPs.** The bread wheat genome sequences and annotation information were downloaded from the EnsemblPlants database ([ftp://ftp.ensemblgenomes.org/pub/plants/release-26/fasta/triticum\\_aestivum/dna/](ftp://ftp.ensemblgenomes.org/pub/plants/release-26/fasta/triticum_aestivum/dna/)). The genomes and gene predictions of the wheat diploid and tetraploid relatives, *Brachypodium distachyon* and *Hordeum vulgare*, were obtained from PGSB PlantsDB (<ftp://ftpmips.helmholtz-muenchen.de/plants/>)<sup>31,56</sup>. First, the keywords “alpha crystallin protein” and “small heat shock protein” were searched against each genome annotation to predict candidate sHSP genes. Then, we used the Hmsearch program in HMMER<sup>57</sup> to search for the family-specific HMM profiles of sHSPs (PF00011) downloaded from the Pfam database (<http://pfam.xfam.org/>)<sup>58</sup>, using each of the predicted proteomes as queries. These two results were merged to remove redundancy and examined for the ACD in the InterPro (<http://www.ebi.ac.uk/interpro/>) and PROSITE (<http://prosite.expasy.org/>) databases. The sequences in which ACD domain were detected in InterPro or PROSITE, were aligned using the M-Coffee program, which combines the output of popular alignment programs<sup>59</sup>. A phylogenetic tree was constructed based on the ACD using the neighbour-joining method and a bootstrap test of 1000 iterations, and the tree was presented by iTOL<sup>60</sup>. Triticeae sHSPs were grouped into different subfamilies by phylogenetic analysis using sHSPs from *A. thaliana* and rice as markers. Wheat sHSPs were mapped to the chromosome based on the annotation information and presented by Circos<sup>61</sup>.

**Molecular evolution analysis.** The codeml program from the PAML4.7 package<sup>46</sup> was used to estimate the  $\omega$  values for Triticeae relatives and sHSP subfamilies. One-ratio model (model 0) which assumes a constant  $\omega$  ratio along all branches and two-ratio model (model 2), which allows different  $\omega$  ratios between foreground and background lineages, were used to detect different selective pressures between the duplicated clades. The site-specific discrete model (model 3), which assumes two classes of sites with different  $\omega$  ratios, and the clade model (model D), which allows selective pressure at one class of sites (foreground clade) to be different from the rest of the phylogeny, were used to detect functional divergence after gene duplication. The branch-site model A (model A), which assumes one class of sites in the foreground lineage  $\omega > 1$  and null model A, with fixed  $\omega = 1$ , were compared to detect positive selection in specific lineages. In PAML analysis, the protein sequence alignments were converted into corresponding codon alignments using the PAL2NAL program<sup>62</sup>. The Selecton program<sup>63</sup> was used to detect the selective pressure acting on each amino acid residue and the SPEER-SERVER<sup>64</sup> was used to identify the type I and type II residues that contribute to functional divergence between duplicated clades.

**Expression profile analysis of wheat sHSPs.** We performed expression profiling analysis of wheat sHSPs using high-throughput RNA-seq data. The RNA-seq data from five tissues (leaf, root, grain, spike and stem) at three development stages were obtained from URGI (<https://urgi.versailles.inra.fr/files/RNASeqWheat/>)<sup>31</sup>, and the RNA-seq data from different cell types of grain at three development stages were downloaded from NCBI (Accession: PRJEB5135)<sup>29</sup>. RNA-seq data from wheat seedlings under heat and drought treatments were downloaded from NCBI (PRJNA257316)<sup>65</sup>. The quality of the public RNA-seq data was checked with the FastQC program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The high-quality paired-end RNA-seq reads from each library were aligned to the wheat reference genome (*Triticum aestivum* IWGSC v2.26; [ftp://ftp.ensemblgenomes.org/pub/plants/release-26/fasta/triticum\\_aestivum/dna/](ftp://ftp.ensemblgenomes.org/pub/plants/release-26/fasta/triticum_aestivum/dna/)) using hisats v2.0.0<sup>66</sup> at “very-sensitive” pre-set parameters in the “end-to-end” mode. The intron length was set to 20 to 5000 nucleotides during alignment. Alignments between reads and wheat reference genome sequences were input into stringTie v1.1.1<sup>67</sup> for the normalization and estimation of the gene expression level. Gene expression levels were reported as fragments per kilobase of exon model per million mapped reads (FPKM). A gene was regarded as expressed in a sample if the FPKM was greater than zero. The wheat genome annotation information used in these analyses was obtained from the EnsemblPlants database ([ftp://ftp.ensemblgenomes.org/pub/plants/release-26/gff3/triticum\\_aestivum](ftp://ftp.ensemblgenomes.org/pub/plants/release-26/gff3/triticum_aestivum)). The expression data were presented by Hemi<sup>68</sup> and iTOL<sup>60</sup>.

**Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) validation.** Total RNA was isolated from the leaves of 7-day seedlings of the hexaploid wheat variety *Chinese Spring*, which had been subjected to heat stress, using a RNA extraction kit with TRIzol reagent (Tiangen Biotech, Beijing, China). A FastQuant RT kit (Tiangen Biotech, Beijing, China) was used to synthesize cDNA from the treated RNA with an oligo-dT primer in a 50- $\mu$ l reaction. Real-time quantitative PCR primers were designed in Primer Premier 5.0 software. The CDS sequence of bread wheat is downloaded from the Ensemble database ([ftp://ftp.ensemblgenomes.org/pub/release-26/plants/fasta/triticum\\_aestivum/cds/](ftp://ftp.ensemblgenomes.org/pub/release-26/plants/fasta/triticum_aestivum/cds/)). The size of the PCR product was set to 100–200 bp and the length of the primer was 15–25 bp. A pre-PCR was performed using wheat cDNA as a template, and primer produced a single specific band was further used for real-time PCR. For PCR, 0.1–0.25  $\mu$ M cDNA were amplified with specific primers for each of eight different gene models, with the 1  $\times$  SYBR Green Master Mix Kit (Applied Biosystems, Foster City, CA, USA) in a final volume of 12.5  $\mu$ L.

## References

- Hua, J. From freezing to scorching, transcriptional responses to temperature variations in plants. *Curr Opin Plant Biol* **12**, 568–573, doi:10.1016/j.pbi.2009.07.012 (2009).
- Mittler, R., Finka, A. & Goloubinoff, P. How do plants feel the heat? *Trends Biochem Sci* **37**, 118–125, doi:10.1016/j.tibs.2011.11.007 (2012).
- Basha, E., O'Neill, H. & Vierling, E. Small heat shock proteins and  $\alpha$ -crystallins: dynamic proteins with flexible functions. *Trends Biochem Sci* **37**, 106–117, doi:10.1016/j.tibs.2011.11.005 (2012).
- Waters, E. R. The evolution, function, structure, and expression of the plant sHSPs. *J Exp Bot* **64**, 391–403, doi:10.1093/jxb/ers355 (2013).
- Kim, K. K., Kim, R. & Kim, S. H. Crystal structure of a small heat-shock protein. *Nature* **394**, 595–599, doi:10.1038/29106 (1998).
- van Montfort, R. L. M., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat Struct Biol* **8**, 1025–1030, doi:10.1038/NSB722 (2001).
- Baldwin, A. J. *et al.* Quaternary Dynamics of alpha B-Crystallin as a Direct Consequence of Localised Tertiary Fluctuations in the C-Terminus. *Journal of molecular biology* **413**, 310–320, doi:10.1016/j.jmb.2011.07.017 (2011).
- Giese, K. C., Basha, E., Catague, B. Y. & Vierling, E. Evidence for an essential function of the N terminus of a small heat shock protein *in vivo*, independent of *in vitro* chaperone activity. *Proc Natl Acad Sci USA* **102**, 18896–18901, doi:10.1073/pnas.0506169103 (2005).
- Jaya, N., Garcia, V. & Vierling, E. Substrate binding site flexibility of the small heat shock protein molecular chaperones. *Proc Natl Acad Sci USA* **106**, 15604–15609, doi:10.1073/pnas.0902177106 (2009).
- Jehle, S. *et al.* N-terminal domain of alphaB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proc Natl Acad Sci USA* **108**, 6409–6414, doi:10.1073/pnas.1014656108 (2011).
- Laganowsky, A. *et al.* Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci* **19**, 1031–1043, doi:10.1002/pro.380 (2010).
- Chauhan, H., Khurana, N., Nijhavan, A., Khurana, J. P. & Khurana, P. The wheat chloroplastic small heat shock protein (sHSP26) is involved in seed maturation and germination and imparts tolerance to heat stress. *Plant Cell Environ* **35**, 1912–1931, doi:10.1111/j.1365-3040.2012.02525.x (2012).
- Kim, K. H. *et al.* Overexpression of a chloroplast-localized small heat shock protein OsHSP26 confers enhanced tolerance against oxidative and heat stresses in tall fescue. *Biotechnol Lett* **34**, 371–377, doi:10.1007/s10529-011-0769-3 (2012).
- Mahesh, U. *et al.* Constitutive overexpression of small HSP24.4 gene in transgenic tomato conferring tolerance to high-temperature stress. *Mol Breeding* **32**, 687–697, doi:10.1007/s11032-013-9901-5 (2013).
- Zhong, L. L. *et al.* Chloroplast Small Heat Shock Protein HSP21 Interacts with Plastid Nucleoid Protein pTAC5 and Is Essential for Chloroplast Development in Arabidopsis under Heat Stress. *The Plant cell* **25**, 2925–2943, doi:10.1105/tpc.113.111229 (2013).
- Hu, X. *et al.* Protein sHSP26 improves chloroplast performance under heat stress by interacting with specific chloroplast proteins in maize (*Zea mays*). *Journal of proteomics* **115**, 81–92, doi:10.1016/j.jpro.2014.12.009 (2015).
- Scharf, K. D., Siddique, M. & Vierling, E. The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing alpha-crystallin domains (Acid proteins). *Cell Stress Chaperon* **6**, 225–237, doi:10.1379/1466-1268(2001)006<0225:TEFOAT>2.0.CO;2 (2001).
- Borges, J. C., Cagliari, T. C. & Ramos, C. H. I. Expression and variability of molecular chaperones in the sugarcane expressome. *J Plant Physiol* **164**, 505–513, doi:10.1016/j.jplph.2006.03.013 (2007).
- Waters, E. R. & Rioflorida, I. Evolutionary analysis of the small heat shock proteins in five complete algal genomes. *J Mol Evol* **65**, 162–174, doi:10.1007/s00239-006-0223-7 (2007).
- Siddique, M., Gernhard, S., von Koskull-Doring, P., Vierling, E. & Scharf, K. D. The plant sHSP superfamily: five new members in Arabidopsis thaliana with unexpected properties. *Cell Stress Chaperon* **13**, 183–197, doi:10.1007/s12192-008-0032-6 (2008).
- Waters, E. R., Aevermann, B. D. & Sanders-Reed, Z. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress Chaperon* **13**, 127–142, doi:10.1007/s12192-008-0023-7 (2008).
- Sarkar, N. K., Kim, Y. K. & Grover, A. Rice sHsp genes: genomic organization and expression profiling under stress and development. *Bmc Genomics* **10**, doi:10.1186/1471-2164-10-393 (2009).

23. Lopes-Caitar, V. S. *et al.* Genome-wide analysis of the Hsp20 gene family in soybean: comprehensive sequence, genomic organization and expression profile analysis under abiotic and biotic stresses. *Bmc Genomics* **14**, doi:10.1186/1471-2164-14-577 (2013).
24. Guo, M. *et al.* Genome-wide analysis of the CaHsp20 gene family in pepper: comprehensive sequence and expression profile analysis under heat stress. *Frontiers in plant science* **6**, 806, doi:10.3389/fpls.2015.00806 (2015).
25. Semenov, M. A. & Shewry, P. R. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Scientific reports* **1**, 66, doi:10.1038/srep00066 (2011).
26. Lobell, D. B. *et al.* The shifting influence of drought and heat stress for crops in northeast Australia. *Global change biology* **21**, 4115–4127, doi:10.1111/gcb.13022 (2015).
27. Lesk, C., Rowhani, P. & Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **529**, 84–87, doi:10.1038/nature16467 (2016).
28. Feldman, M. & Levy, A. A. Genome Evolution Due to Allopolyploidization in Wheat. *Genetics* **192**, 763–774, doi:10.1534/genetics.112.146316 (2012).
29. Pfeifer, M. *et al.* Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091–1250091, doi:10.1126/science.1250091 (2014).
30. Marcussen, T. *et al.* Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**, doi:10.1126/science.1250092 (2014).
31. Mayer, K. F. X. *et al.* A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**, doi:10.1126/science.1251788 (2014).
32. Conant, G. C., Birchler, J. A. & Pires, J. C. Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. *Curr Opin Plant Biol* **19**, 91–98, doi:10.1016/j.pbi.2014.05.008 (2014).
33. Hegarty, M. J. & Hiscock, S. J. Genomic clues to the evolutionary success of review polyploid plants. *Curr Biol* **18**, R435–R444, doi:10.1016/j.cub.2008.03.043 (2008).
34. Wang, J., Tao, F., Marowsky, N. C. & Fan, C. Evolutionary Fates and Dynamic Functionalization of Young Duplicate Genes in Arabidopsis Genomes. *Plant Physiology*, pp. 01177.02016 (2016).
35. Pont, C., Murat, F., Confolent, C., Balzergue, S. & Salse, J. RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (*Triticum aestivum* L.). *Genome biology* **12**, 10.1186/Gb-2011-12-12-R119 (2011).
36. Soltis, P. S., Marchant, D. B., Van de Peer, Y. & Soltis, D. E. Polyploidy and genome evolution in plants. *Current opinion in genetics & development* **35**, 119–125, doi:10.1016/j.gde.2015.11.003 (2015).
37. Pandey, B., Kaur, A., Gupta, O. P., Sharma, I. & Sharma, P. Identification of HSP20 Gene Family in Wheat and Barley and Their Differential Expression Profiling Under Heat Stress. *Appl Biochem Biotech* **175**, 2427–2446, doi:10.1007/s12010-014-1420-2 (2015).
38. Brenchley, R. *et al.* Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491**, 705–710, doi:10.1038/nature11650 (2012).
39. Jia, J. *et al.* Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95, doi:10.1038/nature12028 (2013).
40. Ling, H. Q. *et al.* Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**, 87–90, doi:10.1038/nature11997 (2013).
41. Choulet, F. *et al.* Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**, doi:10.1126/science.1249721 (2014).
42. Dubcovsky, J. & Dvorak, J. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* **316**, 1862–1866, doi:10.1126/science.1143986 (2007).
43. Akpinar, B. A., Lucas, S. J., Vrána, J., Doležel, J. & Budak, H. Sequencing chromosome 5D of *Aegilops tauschii* and comparison with its allopolyploid descendant bread wheat (*Triticum aestivum*). *Plant biotechnology journal* **13**, 740–752, doi:10.1111/pbi.12302 (2015).
44. Zhang, H. *et al.* Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proc Natl Acad Sci USA* **110**, 3447–3452, doi:10.1073/pnas.1300153110 (2013).
45. Waters, E. R., Nguyen, S. L., Eskandar, R., Behan, J. & Sanders-Reed, Z. The recent evolution of a pseudogene: diversity and divergence of a mitochondria-localized small heat shock protein in *Arabidopsis thaliana*. *Genome* **51**, 177–186, doi:10.1139/g07-114 (2008).
46. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–1591, doi:10.1093/molbev/msm088 (2007).
47. Waters, E. R. & Vierling, E. Chloroplast small heat shock proteins: Evidence for atypical evolution of an organelle-localized protein. *P Natl Acad Sci USA* **96**, 14394–14399, doi:10.1073/pnas.96.25.14394 (1999).
48. Waters, E. R. & Vierling, E. The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. *Mol Biol Evol* **16**, 127–139, doi:10.1093/oxfordjournals.molbev.a026033 (1999).
49. Pradhan, G. P. & Prasad, P. V. Evaluation of wheat chromosome translocation lines for high temperature stress tolerance at grain filling stage. *Plos One* **10**, e0116620, doi:10.1371/journal.pone.0116620 (2015).
50. Nussbaumer, T. *et al.* Joint Transcriptomic and Metabolomic Analyses Reveal Changes in the Primary Metabolism and Imbalances in the Subgenome Orchestration in the Bread Wheat Molecular Response to *Fusarium graminearum*. *G3* **5**, 2579–2592, doi:10.1534/g3.115.021550 (2015).
51. Powell, J. J. *et al.* The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias. *Plant biotechnology journal*, doi:10.1111/pbi.12651 (2016).
52. Wang, X. *et al.* Transcriptome asymmetry in synthetic and natural allotetraploid wheats, revealed by RNA-sequencing. *New Phytol* **209**, 1264–1277, doi:10.1111/nph.13678 (2016).
53. ur Rehman, A. *et al.* Screening wheat germplasm for heat tolerance at terminal growth stage. *Plant Omics* **2**, 9–19 (2009).
54. Lobell, D. B. & Tebaldi, C. Getting caught with our plants down: the risks of a global crop yield slowdown from climate trends in the next two decades. *Environ Res Lett* **9**, 074003, doi:10.1088/1748-9326/9/7/074003 (2014).
55. Tack, J., Barkley, A. & Nalley, L. L. Effect of warming temperatures on US wheat yields. *Proc Natl Acad Sci USA* **112**, 6931–6936, doi:10.1073/pnas.1415181112 (2015).
56. Spannagl, M. *et al.* PGSB PlantsDB: updates to the database framework for comparative plant genome research. *Nucleic Acids Res*, doi:10.1093/nar/gkv1130 (2015).
57. Eddy, S. R. A new generation of homology search tools based on probabilistic inference. *Genome informatics. International Conference on Genome Informatics* **23**, 205–211 (2009).
58. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res* **40**, D290–D301, doi:10.1093/nar/gkr1065 (2012).
59. Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **39**, W13–W17, doi:10.1093/nar/gkr245 (2011).
60. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**, W475–W478, doi:10.1093/nar/gkr201 (2011).
61. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639–1645, doi:10.1101/gr.092759.109 (2009).
62. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609–W612, doi:10.1093/nar/gkl315 (2006).

63. Stern, A. *et al.* Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. *Nucleic Acids Res* **35**, W506–W511, doi:[10.1093/nar/gkm382](https://doi.org/10.1093/nar/gkm382) (2007).
64. Chakrabarti, S., Bryant, S. H. & Panchenko, A. R. Functional specificity lies within the properties and evolutionary changes of amino acids. *Journal of molecular biology* **373**, 801–810, doi:[10.1016/j.jmb.2007.08.036](https://doi.org/10.1016/j.jmb.2007.08.036) (2007).
65. Liu, Z. *et al.* Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC plant biology* **15**, 152, doi:[10.1186/s12870-015-0511-8](https://doi.org/10.1186/s12870-015-0511-8) (2015).
66. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357–360, doi:[10.1038/nmeth.3317](https://doi.org/10.1038/nmeth.3317) (2015).
67. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290–295, doi:[10.1038/nbt.3122](https://doi.org/10.1038/nbt.3122) (2015).
68. Deng, W., Wang, Y., Liu, Z., Cheng, H. & Xue, Y. HemI: a toolkit for illustrating heatmaps. *Plos One* **9**, e111988, doi:[10.1371/journal.pone.0111988](https://doi.org/10.1371/journal.pone.0111988) (2014).

## Acknowledgements

X.W. and S.X. acknowledge the financial support of the National Natural Science Foundation of China (31501380), Basic Research Project for Natural Science of Shanxi Province (2016JQ3023), Basic Research Grants for Central Universities (2452015125), grants for PhD Starting Research of Northwest A&F University (109021504) and grants from the National Natural Science Foundation of China (31370318).

## Author Contributions

X.W., Z.L. and S.X. carried out the public genome data collection. X.W., C.M., R.W. and J.C. performed analyses of the data. Q.S., X.W. and S.X. outlined the initial concept of assimilation. X.W., S.X., Q.S., C.M. and Z.W. contributed to the study design. X.W., S.X. and X.S. wrote the manuscript. All authors were involved in the revision of the manuscript and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-01857-3](https://doi.org/10.1038/s41598-017-01857-3)

**Competing Interests:** The authors declare that they have no competing interests.

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



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

© The Author(s) 2017