

# SCIENTIFIC REPORTS

OPEN

## Genome-wide identification and analysis of non-specific Lipid Transfer Proteins in hexaploid wheat

Allan Kouidri<sup>1</sup>, Ryan Whitford<sup>1</sup>, Radoslaw Suchecki<sup>1,2</sup>, Elena Kalashyan<sup>1</sup> & Ute Baumann<sup>1</sup>

Non-specific Lipid Transfer Proteins (nsLTPs) are involved in numerous biological processes. To date, only a fraction of wheat (*Triticum aestivum* L.) nsLTPs (TaLTPs) have been identified, and even fewer have been functionally analysed. In this study, the identification, classification, phylogenetic reconstruction, chromosome distribution, functional annotation and expression profiles of TaLTPs were analysed. 461 putative TaLTPs were identified from the wheat genome and classified into five types (1, 2, C, D and G). Phylogenetic analysis of the TaLTPs along with nsLTPs from *Arabidopsis thaliana* and rice, showed that all five types were shared across species, however, some type 2 TaLTPs formed wheat-specific clades. Gene duplication analysis indicated that tandem duplications contributed to the expansion of this gene family in wheat. Analysis of RNA sequencing data showed that TaLTPs were expressed in most tissues and stages of wheat development. Further, we refined the expression profile of anther-enriched expressed genes, and identified potential *cis*-elements regulating their expression specificity. This analysis provides a valuable resource towards elucidating the function of TaLTP family members during wheat development, aids our understanding of the evolution and expansion of the TaLTP gene family and, additionally, provides new information for developing wheat male-sterile lines with application to hybrid breeding.

Plant non-specific Lipid Transfer Proteins (nsLTPs) are small and soluble proteins with the ability to transfer various lipid molecules between membranes *in vitro*. nsLTPs are characterized by an eight cysteine motif (8 CM) backbone with the general form C-Xn-C-Xn-CC-Xn-CXC-Xn-C-Xn-C<sup>1</sup>. The cysteine residues are linked by four disulphide bonds stabilizing a tertiary structure composed of four or five alpha helices, with a hydrophobic cavity where the lipid binding takes place<sup>2</sup>. In general, nsLTPs possess an N-terminal secretory signal peptide targeting the protein to the secretory pathway. In addition, some LTPs also carry a C-terminal signal sequence whereby a glycosylphosphatidylinositol-anchor (GPI-anchor) is post-translationally attached to the protein; The GPI-anchor tethers the peptide to the extracellular side of the plasma membrane<sup>3</sup>. nsLTPs have been reported to participate in various biological processes, such as plant signalling, plant defence against biotic and abiotic stresses, cuticular wax and cutin synthesis, seed maturation and sexual reproduction<sup>4</sup>.

Plant nsLTPs consist of a large multigene family found in all land plants and are abundantly expressed in most tissues. Initially, nsLTPs were divided into two major groups based on molecular weight of the mature protein: nsLTP1 (9 kDa) and nsLTP2 (7 kDa)<sup>5</sup>. These two groups differ in the disulphide bond linkages, nsLTP1 at Cys1-Cys6 and Cys5-Cys8 and nsLTP2 at Cys1-Cys5 and Cys6-Cys8. Recently, a new classification according to intron position, presence of GPI-anchor pro-peptide domain and amino acid sequence identity was proposed by Edstam *et al.*<sup>6</sup>. The system classified nsLTPs into 10 types, including five major types (Type 1, 2, C, D and G) and five minor types containing fewer members (E, F, H, J and K). nsLTPs have been reported by genome-wide analysis for several members of the *Poaceae*, including maize (*Zea mays*) (63 nsLTPs), rice (*Oriza sativa*)

<sup>1</sup>University of Adelaide, School of Agriculture, Food & Wine, Waite Campus, Urrbrae, South Australia, 5064, Australia.

<sup>2</sup>Commonwealth Scientific and Industrial Research Organization, Agriculture and Food, Waite Campus, Urrbrae, South Australia, 5064, Australia. Correspondence and requests for materials should be addressed to U.B. (email: [ute.baumann@adelaide.edu.au](mailto:ute.baumann@adelaide.edu.au))

	Wheat	Maize <sup>a</sup>	Sorghum <sup>a</sup>	Rice <sup>a</sup>	<i>A. thaliana</i> <sup>a</sup>
Type 1	39	8	9	18	13
Type 2	274	9	7	13	13
Type C	5	2	2	2	3
Type D	57	16	13	14	12
Type E	0	0	0	0	2
Type G	86	26	24	27	29
Single	0	2	3	3	7
Total nsLTPs	461	63	58	77	79

**Table 1.** A summary of nsLTP genes in wheat, maize, sorghum, rice and *A. thaliana*. <sup>a</sup>nsLTPs members retrieved from Wei and Zhong (2014)<sup>7</sup>.

(77 nsLTPs) and sorghum (*Sorghum bicolor*) (58 nsLTPs)<sup>7</sup>. Previously, in wheat 156 nsLTPs were identified based on EST datasets<sup>8</sup>.

Recent studies have revealed the importance of nsLTPs for pollen development. In *A. thaliana*, RNA interference knock-down for *AtLTPg.3* and *AtLTPg.4* displayed deformed or sterile pollen grains<sup>9</sup>. Of the type C nsLTPs, *AtLTPc.1*, *AtLTPc.2* and *AtLTPc.3* have an anther specific expression restricted to the tapetal cell layer<sup>10</sup>. *AtLTPc.3* was shown to be secreted into the anther locule whereby it ultimately becomes a constituent of the microspore surface. Double RNAi silencing of *AtLTPc.1* and *AtLTPc.3* affected intine morphology, however, pollen grains showed no reduction in fertility. Similarly to *A. thaliana* nsLTP genes, the maize (*Zea mays*) *Ms44* also encodes a type C LTP specifically expressed in tapetal cells with its silencing having no effect on fertility<sup>11</sup>. However, a mutation impairing the cleavage of the *Ms44* signal peptide and therefore blocking its secretion, results in a dominant male sterility phenotype. In contrast, silencing of the rice *OsC6*, an anther-specific LTP, resulted in reduced pollen fertility<sup>12</sup>. Different to what has been observed in rice, wheat *TaMs1* is a nsLTP type G, which shows expression specifically in anthers containing pre-meiosis to meiotic microspores<sup>13,14</sup>. Detailed examination of anthers derived from several deletion mutants (*ms1a*, *ms1b*, *ms1c*) and ethyl methanesulfonate (EMS)-derived mutants (*ms1d*, *ms1e*, *ms1f* and *ms1h*) revealed male sterility is a consequence of disrupted orbicule and pollen exine structure. The determination that wheat *TaMs1* is a single locus nuclear-encoded gene necessary for wheat male fertility represented a significant advance towards developing a hybrid wheat production system similar to the maize Seed Production Technology (SPT)<sup>13,15</sup>. In previous studies, only a small portion of nsLTPs from wheat were identified. Considering a wheat genome reference sequence is now approaching completion, an opportunity exists to initiate a genome-wide analysis of the nsLTP gene family for this species.

In this study we identified 461 putative nsLTPs in the bread wheat genome (cv. Chinese Spring). We conducted a comprehensive study on the phylogeny, genomic structure, chromosomal location and expression profiles of the nsLTP gene family in wheat. Our analysis provides new insights into the *TaLTP* gene family which will support future functional research of nsLTPs. We identified anther-enriched nsLTPs of likely involvement in pollen development. When combined with new gene-editing technologies, this opens opportunities for exploring new loci for inducing male sterility that has application to hybrid breeding.

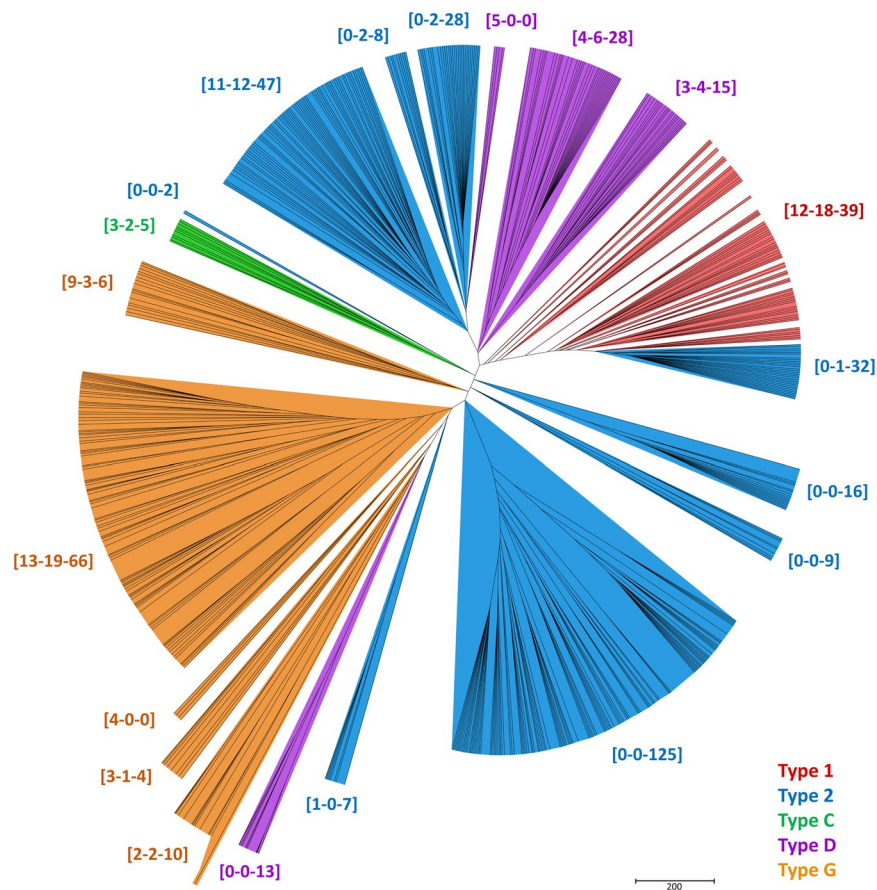
## Results

**Identification and classification of wheat nsLTPs.** A total of 461 putative wheat nsLTPs were identified in cv. Chinese Spring (Supplementary Table S1). Predicted nsLTPs were classified into five types according to Edstam *et al.*<sup>6</sup> (Type 1, 2, C, D and G); type 2 contained most members with 59.44% of wheat nsLTPs, followed by type G (18.66%), type D (12.36%), type 1 (8.46%) and type C (1.08%). The proportion of wheat nsLTPs types varies greatly from that reported in genome-wide analyses performed in *A. thaliana* (*AtLTPs*), rice (*OsLTPs*), maize (*ZmLTPs*) and sorghum (*SbLTPs*) (Table 1), mainly due to a higher proportion of type 2 nsLTPs in wheat.

The evolutionary relationship of nsLTPs between wheat, rice and *A. thaliana* was determined based on phylogenetic analysis (Figs 1 and S1). The tree organisation was in agreement and coherent with the organisation of the five types. For type 1 and type C, all sequences belonging to the same type were grouped and constitute monophyletic groups (i.e. clades). However, type 2, type D and Type G sequences were divided into several groups with five, three and two clades, respectively. In addition, the distribution of nsLTP members within clades was not always quantitatively homogeneous between species, with some clades containing sequences only from wheat or *A. thaliana*. These species-specific clades, such as Type 2 [0-0-9], [0-0-16] and [0-0-125] and type D [5-0-0] and [0-0-13], contained nsLTPs with proline-rich sequences at the N-terminal of the 8 CM.

**Gene and protein structures of the TaLTPs.** Potential protein post-translational modifications of the 461 identified TaLTPs were investigated, including predictions of signal peptide domains and pre-GPI anchor transmembrane domains<sup>16,17</sup>. Following cleavage of the predicted signal peptide and pre-GPI anchor domain, 418 unique mature proteins remained; among these, 386 TaLTPs possessed a unique 8 CM.

nsLTPs are characterized by the highly conserved 8 CM and analysis of the 8 CM consensus within wheat sub-types identified a variable number of inter-cysteine amino acid residues (Table 2). Type 2 TaLTPs contained the most variable spacing across the 8 CM, a likely consequence of the large number of members relative to the other sub-types. All Type C TaLTPs possessed a spacing of 12 residues between the Cys<sub>6</sub> and Cys<sub>7</sub> of the 8 CM, while a spacing of 12 residues was not present in any other TaLTP class.



**Figure 1.** An unrooted phylogenetic tree of nsLTPs from *A. thaliana*, rice and bread wheat. The phylogenetic tree was built from an alignment of the predicted mature proteins. Brackets indicate the number of sequences from *A. thaliana*, rice and wheat, respectively. A detailed phylogenetic tree is provided in Supplementary Fig. S1.

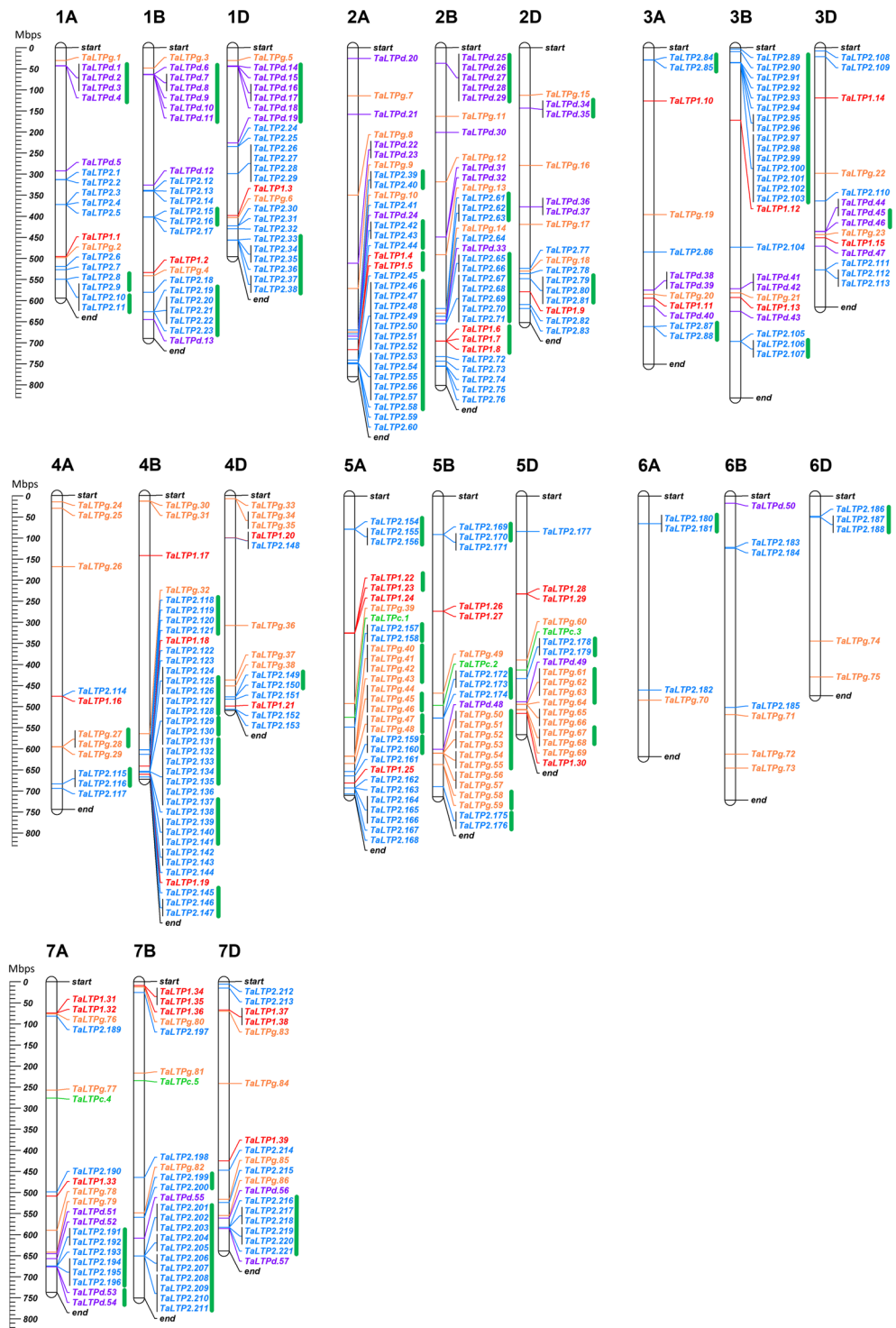
Type	N <sup>o</sup>	Spacing pattern									
1	39	C	X <sub>9</sub>	C	X <sub>13-15,19</sub>	CC	X <sub>19-20,21</sub>	CXC	X <sub>20-24</sub>	C	X <sub>7,13,14,15</sub>
2	274	C	X <sub>3,5,7,8-10,17</sub>	C	X <sub>11,12-19,21,27</sub>	CC	X <sub>8-13,19</sub>	CXC	X <sub>8,10,15,18,19,21-27,36</sub>	C	X <sub>3-9,11,13,15,16,17</sub>
C	5	C	X <sub>9</sub>	C	X <sub>14,19</sub>	CC	X <sub>9</sub>	CXC	X <sub>12</sub>	C	X <sub>6</sub>
D	57	C	X <sub>9,10,12-13,14</sub>	C	X <sub>14,16-18</sub>	CC	X <sub>9-12</sub>	CXC	X <sub>11,17,23,24,27</sub>	C	X <sub>4,7,9,10,11,20</sub>
G	86	C	X <sub>6,9,10,12</sub>	C	X <sub>8,13-18,20</sub>	CC	X <sub>12,14,19</sub>	CXC	X <sub>21-26,29</sub>	C	X <sub>5,6,8-10,13,20</sub>

**Table 2.** Diversity of eight cysteine motifs of TaLTPs. “X” represents any amino number, and the number following “X” indicates the number of amino acid residues. Underlined numbers indicate spacings specific for a type.

When analysing the 8 CM spacing across all TaLTPs, we also identified conservation in the amino acids within these spaces, reflecting higher identity within, but not across sub-types. This was depicted using WebLogo3 tool<sup>18</sup> (Supplementary Fig. S2). For the CXC motif, hydrophobic residues at the X position were observed for most type 2 (86.1%) and type G (87.0%) and type C (100%) proteins, whereas the presence of a hydrophilic residue was predominantly observed in type 1 TaLTPs (69.2%).

To better understand TaLTP protein characteristics, we analysed the isoelectric point (pI) and molecular weight (MW) for all putative TaLTPs. Their MW ranged from 6.73 kDa to 21.73 kDa. Type G nsLTPs have previously been reported to possess the highest MW due to the presence of supernumerary amino acid residues C-terminal to the 8 CM. In contrast, to which has been reported in other species (*Zea mays*, *Marchantia polymorpha*, *Physcomitrella patens*, *Selaginella moellendorffi*, *Adiantum capillus-veneris*)<sup>6,7</sup>, five type D TaLTPs had a higher MW than type G TaLTPs. Furthermore, type 2 nsLTPs previously reported to be 7 kDa proteins, averaged 10.11 kDa in wheat<sup>5</sup>.

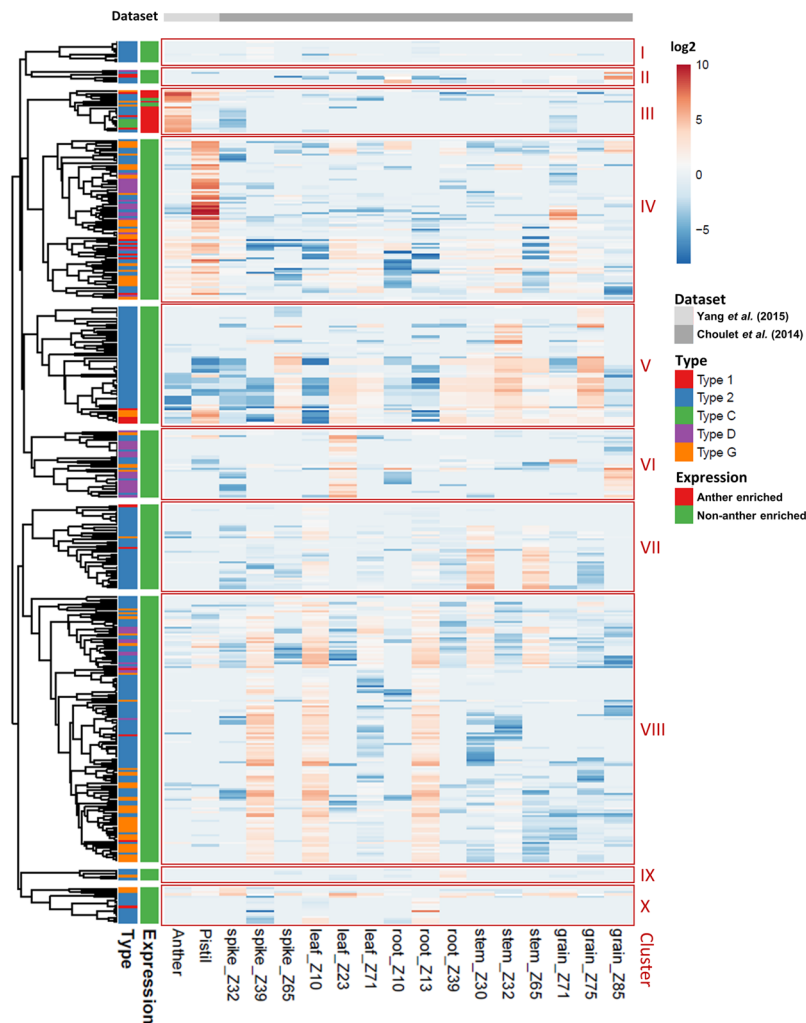
Exon-intron structure was also used to classify wheat nsLTPs, based on criteria proposed by Edstam *et al.*<sup>9</sup>. Accordingly, type 2 TaLTPs were lacking introns, with the exception of 11 genes containing an intron downstream of the 8 CM containing exon and classified as type 2 proteins based on peptide sequence identity (*TaLTP2.71*, *TaLTP2.82*, *TaLTP2.115*, *TaLTP2.127*, *TaLTP132*, *TaLTP2.135*, *TaLTP2.173*, *TaLTP2.217*, *TaLTP2.218*, *TaLTP2.220*



**Figure 2.** Chromosomal location of *TaLTPs*. The scale represents megabases (Mb). Chromosome names are indicated above each vertical bar. Green bars indicate clusters of tandemly duplicated genes.

and *TaLTP2.238*) (Supplementary Fig. S3). All type 1, type C and type D genes contained one intron positioned respectively at nucleotide 5, 1 and 4 after the last cysteine of the 8 CM. In contrast, Type G *TaLTPs* contained up to four introns.

**Chromosomal localization and duplication of *TaLTPs* gene family members.** Of the 461 *TaLTPs*, physical location for 408 *TaLTPs* was identified within the Chinese Spring reference sequence (IWGSC RefSeq v1.0)<sup>19</sup> (Fig. 2). *TaLTPs* were unevenly distributed across the 21 wheat pseudo-molecules. Chromosome 4B contained the highest number of *TaLTPs* (36), while fewest (4) were identified on chromosome 6A. In addition,



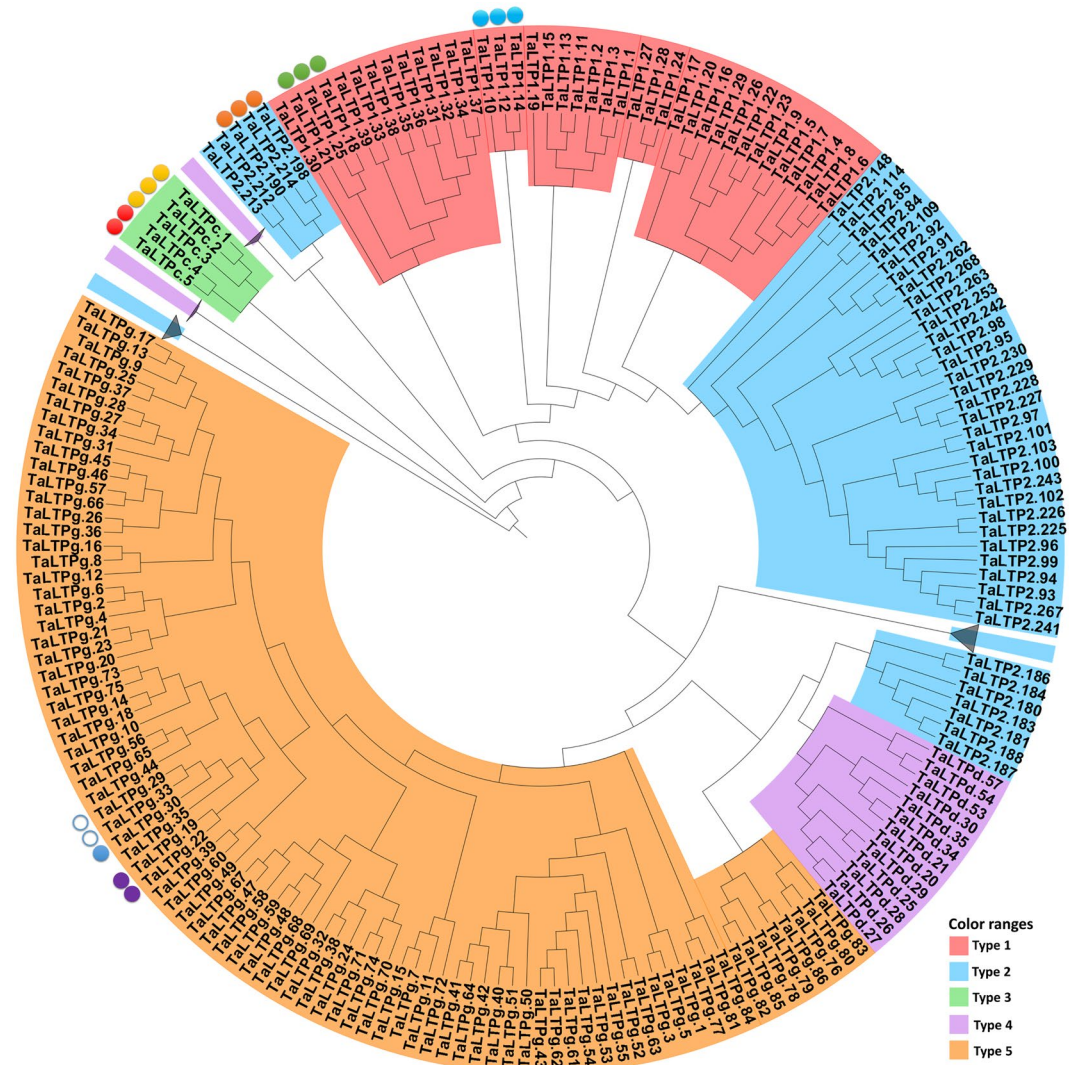
**Figure 3.** Heat map representation and hierarchical clustering of transcript level of *TaLTPs* across different tissues. Data were retrieved from Choulet *et al.* (2014) and Yang *et al.* (2015)<sup>70,71</sup>. Genes were clustered by an average-linkage method (Metsalu, 2005). The transcript abundance in fragments per kilobase of exon per million fragments mapped (FPKM) of each genes is available as Supplementary Table S3.

*TaLTPs* distribution varied across the A (129 *nsLTPs*; 1 *TaLTP*/38.3 Mbps), B (169 *nsLTPs*, 1 *TaLTP*/30.7 Mbps) and D (110 *nsLTPs*; 1 *TaLTP*/35.9 Mbps) sub-genomes.

This uneven density of *TaLTPs* between sub-genomes is a likely consequence of ancestral translocation and duplication events. One such example is the presence of two significant clusters of tandem repeat type 2 *TaLTPs* found on both chromosome 3BS and 4BL relative to their homeologues. We identified 54 tandem duplication clusters involving a total of 200 *TaLTPs*. This suggests that some *TaLTPs* have undergone more than one round of duplication. Among the 200 duplicated genes, 144 belong to type 2 *TaLTPs*. This may explain why type 2 is found to be over-represented in the wheat genome relative to other species.

To analyse the evolutionary processes of type 2 *TaLTPs*, we performed non-synonymous ( $K_a$ ) and synonymous ( $K_s$ ) substitution ratio analysis of the three largest *TaLTP* duplication clusters (Chr2A, *TaLTP*2.46-*TaLTP*58; Chr3B, *TaLTP*2.91-*TaLTP*2.103 and Chr7B, *TaLTP*2.201-*TaLTP*2.211) (Supplemental Table S2; Fig. 2). It was found that most duplicated pairs had a ratio  $K_a/K_s$  lower than one, implying that the genes evolved under influence of purifying selection with limited functional divergence after duplication. The  $K_a/K_s$  of *TaLTP*2.54/*TaLTP*2.49 gene pair was 1.0805, indicating a neutral selection pressure in evolution. Furthermore, divergence time of the duplicated genes were estimated by based on the number of non-synonymous substitution ( $K_s$ ). Out of the 31 duplicated pairs analysed the  $K_s$  values were null for 38.7% of them, indicating recent duplication events. For the remaining duplicated pairs, their corresponding duplication age were estimated to vary from 0.73 million years ago (MYA) to 53.9 MYA.

**Expression analysis of *TaLTPs*.** The expression profiles of all 461 predicted *TaLTPs* were first analysed using public RNA-seq datasets from seven different tissues including leaves, roots, grains, stems, spikes, pistils and anthers (Fig. 3, Supplementary Table S3). Their expression patterns varied greatly across tissues and



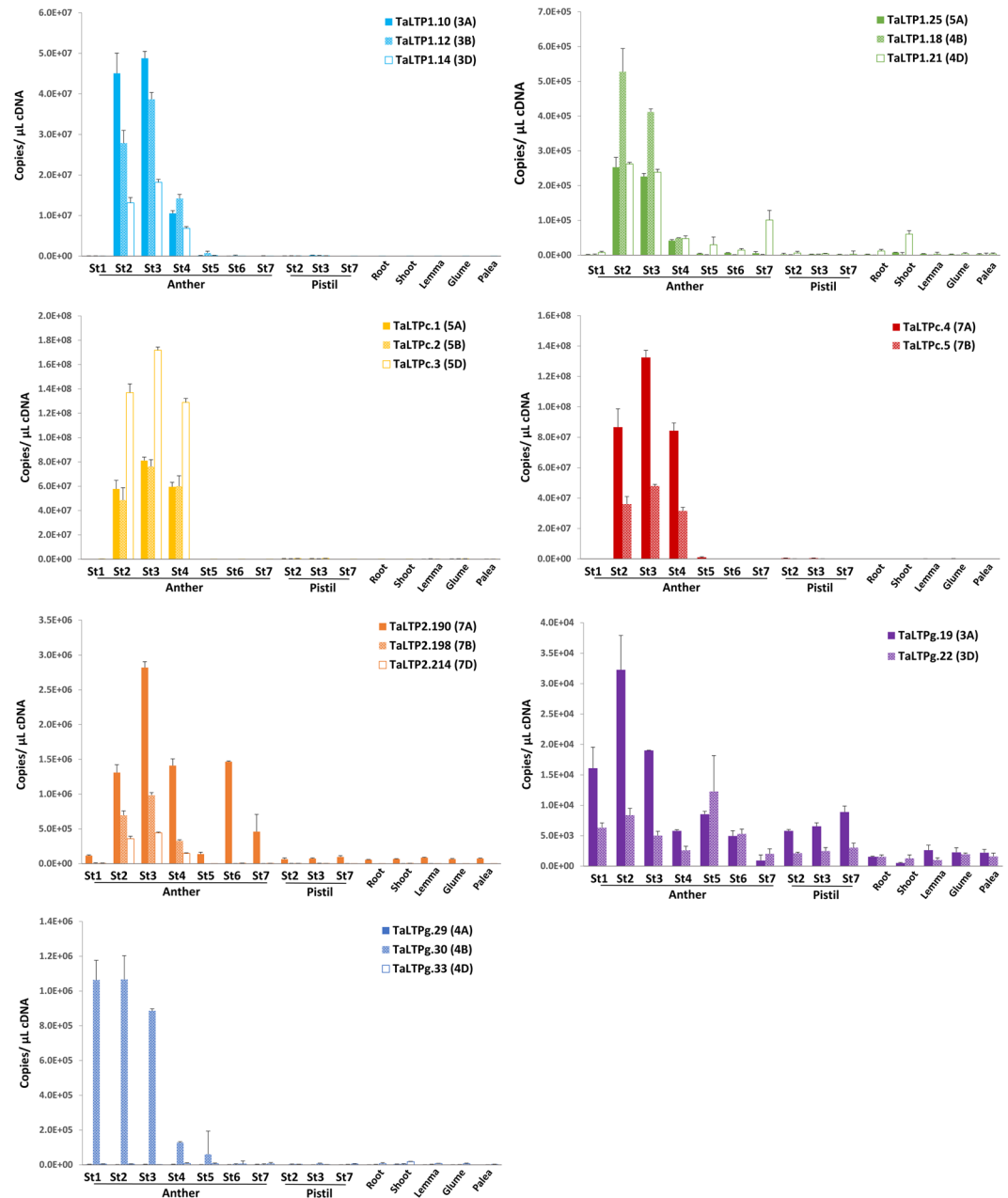
**Figure 4.** Unrooted phylogenetic tree of *TaLTPs* highlighting potential anther-enriched genes. Filled circles represent anther-enriched genes identified by RNA-seq<sup>70,71</sup>, empty circles represent homologous genes with no expression detected; different colors were used for different homeoloci. For more visibility, subtrees including sequences of the same type are grouped and represented by a grey triangle. The fully extended tree is available as Supplementary Fig. S4. The phylogenetic tree was built from an alignment of the predicted mature proteins.

developmental stages. No expression was detected for 30 *TaLTPs*, indicating that they are either pseudogenes or expressed in tissues or under specific environmental conditions for which RNA-seq data was not available.

Based on expression profiles, *TaLTPs* were divided into ten clusters (I to X) by hierarchical ordering (i.e. Cluster III contains anther-enriched genes while Cluster IV mostly contains genes with highly expressed in pistils). Type 1, type 2 and type G *TaLTPs* were present in most clusters, reflecting ubiquitous expression across tissues and developmental stages. In contrast, type C *TaLTPs*, only present in Cluster III, are found preferentially expressed in anthers. Comparing the proportion of *TaLTPs* per type and the proportion of *TaLTP* types per cluster found no significant distribution differences within Cluster VIII. In contrast, Cluster V (32 members) contained genes with expression enriched in leaf (Z23) and grain (Z85), and accounted for 56.2% of type D *TaLTPs* while overall, type D *TaLTPs* comprise only 12.4% of total *TaLTPs*.

In order to identify genes potentially involved in pollen development, we focussed on *TaLTPs* genes preferentially expressed in anthers. In total, 17 *TaLTPs* showed an anther-enriched expression profile based on the RNA-seq data. Among these, only *TaMs1*, a type G *nsLTP*, has been reported as anther-specific, and demonstrated to be involved in pollen exine development<sup>13,14</sup>. Two of the loci containing putative anther-expressed genes were found to possess only two *nsLTPs* homeologues from the three sub-genomes (*TaLTPc4/TaLTPc.5* and *TaLTPg.19/TaLTPg.22*).

The evolutionary relationship of all identified anther-expressed genes was demonstrated using phylogenetic analysis (Fig. 4). The selected *TaLTPs* were present within different clades, suggesting that these genes are derived from different ancestors. For the purpose of validating their anther-enriched expression profile, and to obtain



**Figure 5.** Expression levels of anther-enriched expressed TaLTP candidates. Each graph represents the expression profile for one homeologue group. St1, archesporial cells; St2, pre-meiotic pollen mother cells; St3, meiotic microspores; St4, early uninucleate; St5, late uninucleate; St6, binucleate; St7, mature pollen. Error bars reflect standard error of three independent tissue replicates ( $n = 3$ ).

more precise information on their expression timing during male gametogenesis, we conducted qRT-PCR across eight different wheat tissues including leaves, shoot, roots, glumes, lemmas, paleas, ovaries and anthers.

qRT-PCR results confirmed that the expression profiles for all selected genes were in agreement with the RNA-seq data (Fig. 5). These anther-enriched genes were highly up-regulated in anthers containing meiotic microspores, with the exception of TaLTPg.30 (*TaMs1*) which exhibited expression in anthers deemed to contain microspores at pre-meiosis. In addition, TaLTPg.30 was the only gene to show expression on only one sub-genome, whereas all other anther-enriched LTPs showed expression across two or all three sub-genomes.

**Promoter analysis of anther-specific TaLTPs.** To evaluate the presence of *cis*-elements within *TaLTP* promoter regions involved in anther-specific expression, we searched for over-represented motifs in anther-specific vs non-anther-specific sequences using MEME suite<sup>20</sup>. First, we focussed on nine boxes deemed to be associated with anther-enriched expression (Supplementary Table S4). Among these boxes, only the element POLLENILELAT52 (AGAAA) was identified to be enriched in anther-specific promoter

Motif	Enrichment		Related	Logo	RC Logo
	<i>P</i> -value <sup>a</sup>				
1	5.80E-10		N/A		
2	3.90E-08		N/A		
3	1.60E-07		N/A		
4	3.70E-07		SMZ (AP2-ERF)		
5	8.80E-07		N/A		
6	1.20E-06		bZIP190 (Zipper-type)		

**Figure 6.** List of motifs enriched in anther-specific nsLTPs promoter regions. The enrichment *P*-value of the motif that are enriched in 17 anther-specific gene promoter regions compared with promoter sequences of the remaining 444 identified TaLTPs. Related, annotation of the motif; Logo, 5′–3′ sequence; RC Logo, reverse-complement sequence; AP2-ERF, APETALA2-ethylene response factor; SMZ, SCHLAFMÜTZE.

regions ( $P$ -value =  $6.72e^{-3}$ )<sup>21</sup>. Secondly, we searched DNA motifs without *a priori* which were enriched in the anther-specific *TaLTPs* promoters relative to the remaining promoter sequence. A total of 6 motifs were identified to be significantly enriched in the anther-specific promoters (Fig. 6). Among these, two motifs were retrieved to be transcription factor binding sites (TFBSs); the TCTCGTAT motif (4), a putative binding site of APETALA2-ethylene response factor (AP2-ERF)<sup>22</sup> and the ACGT core motif (6), a potential bZIP binding site<sup>23</sup>.

## Discussion

In the current study, a total of 461 *nsLTPs* were identified in the wheat genome (cv. Chinese Spring), including 39 type 1, 274 type 2, five type C, 57 type D and 86 type G. In comparison to *A. thaliana*, rice and maize, we found the wheat genome to contain over 5 times the number of *nsLTPs* (Table 1). The expansion of wheat *nsLTPs* could be due to the following reasons: (i) bread wheat is an allohexaploid species that originated from hybridization events involving three different diploid progenitors (AABBDD)<sup>24</sup> (ii) small-scale gene duplication, including segmental and tandem duplication, may have played a significant role in *nsLTPs* gene family evolution in wheat. Genome duplication is generally accepted to be a primary source of genetic novelty through subfunctionalization and neofunctionalization and has been central to the evolution of angiosperms, leading to species divergence<sup>25</sup>. Duplication events of *nsLTPs* have been reported in various species, such as *A. thaliana*, turnip, rice, maize and cotton<sup>7,8,26,27</sup>. In this study, phylogenetic relationships between wheat, rice and *A. thaliana* *nsLTPs* showed an expansion of type 2 genes in wheat relative to rice and *A. thaliana* (Fig. 1). This is reflected by the presence of type 2 clades containing only wheat sequences, mostly present in tandem repeats on the wheat pseudomolecules (Fig. 2, Supplemental Table S2). We identified 54 tandem duplication clusters accounting for 200 *TaLTPs*, 72% of which were classified as type 2. Genic duplication and functional redundancy allows these gene sequences to accumulate mutations, increasing divergence and, over time, leading to expansion and evolution of the gene family<sup>28</sup>. (iii) Natural selection would favour duplication of genes that results in adaptive expansion of gene families, it is suggested that retention and expansion of resistance genes in bread wheat might be the results of selection during domestication<sup>29</sup>. Similar expansion was reported for the resistance *NBS-LRR* gene family in wheat<sup>30</sup>.

Phylogenetic relationships between *A. thaliana*, rice and wheat *nsLTPs* showed that some clades contained only wheat sequences (Fig. 1). These wheat specific clades were enriched in *TaLTP* sequences with high proline content at the N-terminus of the 8 CM. Here, we identify 142 proline rich *TaLTPs* (Supplemental Table S7). It is worth noting that *nsLTPs* from rice and Arabidopsis listed by Wei and Zhong (2014)<sup>7</sup> did not incorporate such hybrid proline-rich proteins (HyPRPs) in their analysis. This could explain the presence of some wheat-specific *nsLTP* clades in our analysis. HyPRPs are suggested to be putative cell wall proteins<sup>31,32</sup> that are typically responsive to multiple biotic and abiotic stress factors<sup>33–35</sup>. The hydrophobic 8 CM domain coupled to the hydrophilic proline-rich domain within the same polypeptide is indicative of a role for such proteins to function at the interface between the hydrophobic plasma membrane and the hydrophilic cell wall, and are reported to be involved in plant cell elongation<sup>32</sup>.

The observed expansion of *nsLTP* gene family in wheat relative to rice and the recent divergence time of the analysed *TaLTP* duplication clusters, which averaged 5.28 MYA (Supplemental Table S2), is in accordance with



the divergence time of rice from *Triticeae* estimated of 50 MYA<sup>36</sup>. With the near complete genome assemblies of some *Triticeae* subgroups including, *Hordeum vulgare*, *Triticum urartu*, *Aegilops tauschii* and *T. turgidum ssp dicoccoides*<sup>37–40</sup>, new opportunities arise for a more thorough evolutionary analysis of the *nsLTP* gene family and further investigation of the duplication events that underpin the gene family expansion in wheat.

*nsLTPs* belong to a large family of pathogenesis-related proteins (PRPs), and are reported to play a role in defence against bacterial and fungal pathogens<sup>41</sup>. Some wheat *nsLTPs* have been shown to have an anti-fungal activity toward wheat and non-wheat fungal pathogens *in vitro* (TaLTP2.99, TaLTP2.241, TaLTP2.267, TaLTP2.227, TaLTP2.228, TaLTP2.229, TaLTP2.230, TaLTP2.242, TaLTP2.253, TaLTP2.263, TaLTP2.268, TaLTP2.109, TaLTP2.91, TaLTP2.92, TaLTP2.262, TaLTP1.22, TaLTP1.23)<sup>42</sup>. Surprisingly, no correlation was observed between their ability to inhibit pathogenic growth and lipid binding activity. However, it has been suggested that their toxicity could be derived from an alteration of the fungal membrane permeability<sup>42</sup>. Additionally, the wheat TaLTP1.22 (100% similarity with the tandem duplicate TaLTP1.23) was reported to be associated with resistance against *Fusarium graminearum*, and its transcript level was at least 50-fold more abundant in plants carrying the resistant allele Qfhs.ifa-5A<sup>43</sup>. Interestingly, these previously reported *TaLTP* genes involved in abiotic stress resistance from type 1 and type 2, were all retrieved within the same phylogenetic clades (Supplemental Fig. S4). In addition, these *TaLTPs* demonstrated a similar expression profile based on RNA-seq analysis, and grouped within the same cluster (V), showing high expression in stems (Z30, Z32 and Z65), leaves (Z23 and Z71) and grains (Z65). Therefore, we speculate that these *TaLTPs* may also possess antifungal activities. In rice, similar inhibition tests using LTP110 demonstrated a critical role for the residues Tyr17 and Arg46 and Pro72 in antifungal activity<sup>44</sup>. As expected, these residues were identified as highly conserved in type 1 *TaLTPs* (Tyr14 and Arg50 and Pro78) (Supplementary Fig. S2).

In diverse crops, the use of male sterility has been exploited for production of hybrid varieties that capture the benefits of hybrid vigour. It is reported that wheat hybrid vigour offers a yield gain of over 10% and improved yield stability<sup>45</sup>. One major limitation towards developing a commercially viable hybrid seed production platform in wheat was the lack of identification of a single locus genic male-sterile mutant and its associated wild-type restorer sequence<sup>15,46</sup>. This limitation has been overcome, only recently, by the identification of the dominant fertility gene *TaMs1*, and the dominant sterility gene *Ms2*<sup>13,14,47,48</sup>. In addition, the recent development of novel gene-editing technologies opens opportunities to generate loss-of-function mutants in a single transgenic event in wheat<sup>49</sup>. *nsLTPs* represent potential candidates towards developing new male sterile mutants as studies have shown that defects in certain anther-expressed *nsLTPs* result in male sterility<sup>10,11,13,50</sup>. In order to identify *TaLTPs* responsible for ensuring male fertility, we analysed the expression profiles of selected *TaLTPs* using RNA-seq data and showed that most *TaLTPs* were expressed across a range of tissues and developmental stages. Interestingly, some members exhibited tissue-specific expression, suggesting important roles in physiological processes during wheat development. This included, 16 *TaLTP* genes (7 loci) which were preferentially or specifically expressed in anthers. These include, three type 2 (single locus), six type 1 (two loci), five type C (two loci) and 2 type G (two loci) genes. Orthologues for these wheat anther-expressed genes have been analysed in rice, maize and sorghum (Supplementary Table S5). Among the identified putative orthologs, only the maize type C *nsLTP* (*Ms44*) has been reported to be involved in male fertility<sup>11</sup>. In addition, no orthologues for *TaLTP2.190*, *TaLTP2.198* and *TaLTP2.214* could be retrieved for the studied species, suggesting a gain of function in wheat for male reproductive organ development. qRT-PCR analysis confirmed the expression profiles of these anther-expressed *TaLTPs* (Fig. 5). Among the anther-expressed *TaLTPs*, only *TaMs1* was identified as a single homeologue expressed gene.

To understand the promoter specificity of the identified anther-specific *TaLTPs* and to also identify putative *cis*-elements controlling their spatial and temporal expression, we identified enriched DNA motifs on promoter regions for the anther-specific *TaLTPs* relative to non-anther enriched *TaLTPs* (Supplementary Table S4, and Fig. 6). A total of seven motifs were identified including three previously reported elements. (i) The first is POLLENLELAT52 (AGAAA), an enhancer element of LAT52 deemed to be essential for high level of expression in pollen (Supplementary Table S4, and Fig. 6)<sup>51</sup>. (ii) The second is an AP2/ERF binding site element identified to regulate SMZ, involved in regulation of floral development in *A. thaliana*<sup>22</sup>. (iii) The third is a binding site element of bZIP190 (ACGT core), reported to be preferentially expressed in flowers of *Antirrhinum majus*<sup>23</sup>. An analysis by truncated transcriptional reporter fusions of the promoter would help to confirm functionality of these identified *cis*-elements in anther-enriched expression.

In conclusion, this is the first comprehensive and systematic analysis of *nsLTPs* in wheat. The structure, classification, evolution and expression profiles of 461 putative *TaLTPs* were analysed. The results of this study revealed the ubiquitous expression of *TaLTPs* during growth and development. The expansion of *TaLTPs* in the wheat genome was attributed to duplications during evolution. Our expression analysis may provide a solid basis for future studies of *TaLTP* function during wheat development. The identification of anther-expressed *TaLTPs* opens opportunities for the development of new male-sterile wheat lines for hybrid seed production.

## Materials and Methods

**Sequence retrieval and structural analysis.** All *nsLTP* sequences of *Arabidopsis thaliana*, rice (*Oriza sativa*), sorghum (*Sorghum bicolor*) and maize (*Zea mays*) previously identified by Wei and Zhong (2014) were retrieved from phytozome<sup>7,52</sup>. To identify wheat putative *nsLTPs* we first used rice protein sequences (LOC\_Os10g36070.1, LOC\_Os07g18750.1, LOC\_Os05g47700.1, LOC\_Os05g40010.1, LOC\_Os03g07100.1, LOC\_Os01g68580.1, LOC\_Os01g59870.1, LOC\_Os01g12020.1) as queries to search against the Wheat IWGSC RefSeq v1.0<sup>19</sup> using tBLASTn with an *e*-value of  $\leq 10$ . Secondly, we examined hit sequences (+/−1000 bps) in all six DNA frames for presence of the 8 CM, and all hits lacking the essential Cys residues were excluded. Then, selected genomic sequences were used for gene prediction using TriAnnot and FGENESH programs with default parameters<sup>53,54</sup>. Subsequently, all predicted genes lacking the 8 CM were removed from further analysis. The remaining candidates were submitted to SUPERFAMILY 2 Beta tool, sequences annotated as proteinase/alpha-amylase

inhibitor and seed storage family, 2S albumin were discarded (<http://beta.supfam.org/>). Additionally, putative nsLTPs lacking NSSs (examined with SignalP 4 server<sup>55</sup>) were also excluded. Presence of a C-terminal GPI-anchor signal was predicted using three prediction tools PredGPI, big PI-plant predictor and GPI-SOM<sup>17,56,57</sup>. Splice junctions were predicted using Splign Transcript to Genomic Alignment tool<sup>58</sup>. The nsLTP nomenclature was based on guidelines from Edstam *et al.*<sup>6</sup>.

**Phylogenetic analysis and gene duplications.** nsLTPs sequences ID of rice (*Oriza sativa*), Sorghum (*Sorghum bicolor*), and Maize (*Zea Mays*) were retrieved from Wei and Zhong<sup>7</sup>. nsLTPs amino acid sequences were downloaded from ensembl plant<sup>59</sup>. All nsLTPs 8 CM sequences of the mature proteins were aligned using Clustal Omega<sup>60</sup>. After manually refinement of the multiple sequence alignment, the phylogenetic tree was built from alignment of the predicted mature proteins by Unweighted Pair Group Method with Arithmetic mean method using the Jalview program<sup>61</sup>. Trees were visualized using the iTOL web tool V4.0.3<sup>62</sup>.

**Gene structure analysis.** In order to identify the precise splice-junction site of predicted *TaLTPs*, coding sequences (CDS) were aligned to genomic sequence using Splign alignment tool<sup>58</sup>. Schematic representation of *TaLTPs* gene structures was generated using the GSGS 2.0 tool<sup>63</sup>.

**Chromosomal mapping and duplication.** *TaLTPs* were mapped onto the 21 wheat chromosomes according to their physical position (bp), from the short arm telomere to the long arm telomere based on IWGSC RefSeq V1.0<sup>19</sup>. MapChart was used to draw their location onto the physical map of each chromosome<sup>64</sup>. To detect gene duplications, the CDS sequences of *TaLTPs* in wheat were blasted against each other and selected with the following cut-offs: 80% of coverage with the similarity of the aligned regions above 80%. Tandemly duplicated *TaLTPs* were defined as two or more adjacent homologous genes located on a single chromosome. The estimation of the evolution rates of the duplicated *TaLTPs* was calculated using KaKs\_calculator 2.0<sup>65</sup>. The divergence times (T) of *TaLTP* duplicated pairs were calculated as  $T = Ks/2r \times 10^{-6}$  Mya, with a divergence rate (r) of  $6.5 \times 10^{-9}$ <sup>66</sup>.

**Gene expression profiles based on RNA-seq.** Previous wheat cv. Chinese Spring transcriptome examinations were performed by RNA-seq to investigate gene expression pattern of genes during vegetative and reproductive development. The transcript abundance of each gene was estimated by fragments per kilobase of exon per million fragments mapped (FPKM). Expression data of *TaLTPs* in spikes, leaves, roots, stems and grains, each sampled at three developmental stages, were retrieved from RNA-seq data downloaded from the European Nucleotide Archive database with the study number PRJEB5314 (<http://www.ebi.ac.uk/ena/data/view/PRJEB5314>). *TaLTPs* expressions data in pistils and anthers were downloaded from NCBI-SRA database with accession number SRP038912.

Heatmaps were generated from log<sub>2</sub>-transformed FPKM using the ClustVis web tool<sup>67</sup>.

**Quantitative RT-PCR analysis.** Total RNA was isolated using ISOLATE II RNA Mini Kit (Biolone, Sydney, Australia) from wheat of cv. Chris tissues including roots, shoot apical meristem (SAM) and glume, lemma, palea, ovaries, and anthers containing microspores from pre-meiosis to maturity. Microspores were cytologically examined for stage of development. The remaining two anthers from the same floret were isolated and snap frozen in liquid nitrogen. All total RNA samples were treated with DNase I (Qiagen). 0.6 µg of RNA was used to synthesise the oligo (dT)-primed first strand cDNA using the superscript IV reverse transcriptase (Thermo Fisher, Adelaide, Australia). Quantitative real-time PCR was performed according to Burton *et al.*<sup>68</sup> using the primer combinations shown in Supplemental Table S6. Amplification products from qRT-PCR on each tissue sample, three technical replicates and three biological replicates were used to estimate the transcript abundance of genes of interest relative to TaActin, TaGAPdH and Ta13-3-3 reference transcripts.

**Promoter analysis.** For motif enrichment analysis using known *cis*-element deemed to be involved in pollen enriched expression (Supplementary Table S4) we used the AME web tool implemented in MEME suite<sup>20</sup> with the following parameter: Ranksum enrichment test with a *P*-value threshold of 0.05.

For *de novo* motif discovery and enrichment, we used the DREME web tool implemented in MEME suite (version 4.12.0)<sup>20</sup> with an e-value threshold of 0.05. MEME was also used to generate sequence logos for each discovered motif. Jaspas database was searched to find related transcription factor binding sites to the motifs identified using a relative profile score threshold of 80%<sup>69</sup>.

## Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## References

- Salminen, T. A., Blomqvist, K. & Edqvist, J. Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta* **244**, 971–997 (2016).
- Finkina, E. I., Melnikova, D. N., Bogdanov, I. V. & Ovchinnikova, T. V. Lipid transfer proteins as components of the plant innate immune system: Structure, functions, and applications. *Acta Naturae* **8**, 47–61 (2016).
- Mayor, S. & Riezman, H. Sorting GPI-anchored proteins. *Nat. Rev. Mol. Cell Biol.* **5**, 110–20 (2004).
- Carvalho, A., de, O. & Gomes, V. M. Role of plant lipid transfer proteins in plant cell physiology-A concise review. *Peptides* **28**, 1144–1153 (2007).
- Dougliez, J. P., Michon, T., Elmorjani, K. & Marion, D. Structure, biological and technological functions of lipid transfer proteins and indolines, the major lipid binding proteins from cereal kernels. *J. Cereal Sci.* **32**, 1–20 (2000).
- Edstam, M. M., Viitanen, L., Salminen, T. A. & Edqvist, J. Evolutionary history of the non-specific lipid transfer proteins. *Mol. Plant* **4**, 947–964 (2011).
- Wei, K. & Zhong, X. Non-specific lipid transfer proteins in maize. *BMC Plant Biol.* **14**, 281 (2014).

8. Boutrot, F., Chantret, N. & Gautier, M.-F. Genome-wide analysis of the rice and Arabidopsis non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. *BMC Genomics* **9**, 86 (2008).
9. Edstam, M. M. & Edqvist, J. Involvement of GPI-anchored lipid transfer proteins in the development of seed coats and pollen in Arabidopsis thaliana. *Physiol. Plant.* **152**, 32–42 (2014).
10. Huang, M.-D., Chen, T.-L. L. & Huang, A. H. C. Abundant Type III Lipid Transfer Proteins in Arabidopsis Tapetum Are Secreted to the Locule and Become a Constituent of the Pollen Exine. *Plant Physiol.* **163**, 1218–1229 (2013).
11. Fox, T. *et al.* A single point mutation in *Ms44* results in dominant male sterility and improves nitrogen use efficiency in maize. *Plant Biotechnol. J.* 1–11, <https://doi.org/10.1111/pbi.12689> (2017).
12. Zhang, D. *et al.* OsC6, encoding a lipid transfer protein, is required for postmeiotic anther development in rice. *Plant Physiol.* **154**, 149–62 (2010).
13. Tucker, E. J. *et al.* Molecular identification of the wheat male fertility gene *Ms1* and its prospects for hybrid breeding. *Nat. Commun.* **8**, 869 (2017).
14. Wang, Z. *et al.* Poaceae-specific *MS1* encodes a phospholipid-binding protein for male fertility in bread wheat. *Proc. Natl. Acad. Sci.* 201715570, <https://doi.org/10.1073/pnas.1715570114> (2017).
15. Whitford, R. *et al.* Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. *J. Exp. Bot.* **64**, 5411–28 (2013).
16. Nielsen, H. *Protein Function Prediction.* **1611**, 59–73 (2017).
17. Pierleoni, A., Martelli, P. & Casadio, R. PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* **9**, 392 (2008).
18. Crooks, G., Hon, G., Chandonia, J. & Brenner, S. NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res* **14**, 1188–1190 (2004).
19. International Wheat Genome Sequencing Consortium. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science (80-.)*. **361** (2018).
20. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, 202–208 (2009).
21. Zhou, D. X. Regulatory mechanism of plant gene transcription by GT-elements and GT-factors. *Trends Plant Sci.* **4**, 210–214 (1999).
22. Wang, P. *et al.* Expansion and Functional Divergence of AP2 Group Genes in Spermatophytes Determined by Molecular Evolution and Arabidopsis Mutant Analysis. *Front. Plant Sci.* **7**, 1–15 (2016).
23. Martínez-García, J. F., Moyano, E., Alcocer, M. J. C. & Martín, C. Two bZIP proteins from Antirrhinum flowers preferentially bind a hybrid C-box/G-box motif and help to define a new sub-family of bZIP transcription factors. *Plant J.* **13**, 489–505 (1998).
24. Feldman, M. & Levy, A. A. Genome evolution due to allopolyploidization in wheat. *Genetics* **192**, 763–774 (2012).
25. Davies, T. J. *et al.* Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proc. Natl. Acad. Sci.* **101**, 1904–1909 (2004).
26. Li, J. *et al.* Genome-wide survey and expression analysis of the putative non-specific lipid transfer proteins in Brassica rapa L. *PLoS One* **9** (2014).
27. Li, F. *et al.* Genomic Identification and Comparative Expansion Analysis of the Non-Specific Lipid Transfer Protein Gene Family in Gossypium. *Sci. Rep.* **6**, 38948 (2016).
28. Conant, G. C. & Wolfe, K. H. Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* **9**, 938–950 (2008).
29. Demuth, J. P. & Hahn, M. W. The life and death of gene families. *BioEssays* **31**, 29–39 (2009).
30. Gu, L., Si, W., Zhao, L., Yang, S. & Zhang, X. Dynamic evolution of NBS-LRR genes in bread wheat and its progenitors. *Mol. Genet. Genomics* **290**, 727–738 (2015).
31. Dvořáková, L., Cvrčková, F. & Fischer, L. Analysis of the hybrid proline-rich protein families from seven plant species suggests rapid diversification of their sequences and expression patterns. *BMC Genomics* **8**, 1–16 (2007).
32. Dvořáková, L., Srba, M., Opatrný, Z. & Fischer, L. Hybrid proline-rich proteins: Novel players in plant cell elongation? *Ann. Bot.* **109**, 453–462 (2012).
33. Bouton, S., Viau, L., Lelièvre, E. & Limami, A. M. A gene encoding a protein with a proline-rich domain (MtPPRD1), revealed by suppressive subtractive hybridization (SSH), is specifically expressed in the Medicago truncatula embryo axis during germination. *J. Exp. Bot.* **56**, 825–832 (2005).
34. Zhang, Y. & Schläppi, M. Cold responsive EARLII type HyPRPs improve freezing survival of yeast cells and form higher order complexes in plants. *Planta* **227**, 233–243 (2007).
35. Priyanka, B., Sekhar, K., Reddy, V. D. & Rao, K. V. Expression of pigeonpea hybrid-proline-rich protein encoding gene (CchYPRP) in yeast and Arabidopsis affords multiple abiotic stress tolerance. *Plant Biotechnol. J.* **8**, 76–87 (2010).
36. Chalupska, D. *et al.* Acc homoeoloci and the evolution of wheat genomes. *Proc. Natl. Acad. Sci. USA* **105**, 9691–6 (2008).
37. Luo, M. C. *et al.* Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. *Nature* **551**, 498–502 (2017).
38. Beier, S. *et al.* Construction of a map-based reference genome sequence for barley, Hordeum vulgare L. *Sci. Data* **4**, 170044 (2017).
39. Ling, H. Q. *et al.* Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature* **496**, 87–90 (2013).
40. Avni, R. *et al.* Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science (80-.)*. **357**, 93–97 (2017).
41. Sudisha, J., Sharathchandra, R. G., Amruthesh, K. N., Kumar, A. & Shetty, H. S. Plant Defence: Biological Control. 379–403, <https://doi.org/10.1007/978-94-007-1933-0> (2012).
42. Sun, J.-Y. *et al.* Characterization and Antifungal Properties of Wheat Nonspecific Lipid Transfer Proteins. *Mol. Plant-Microbe Interact.* **21**, 346–360 (2008).
43. Schweiger, W. *et al.* Transcriptomic characterization of two major Fusarium resistance quantitative trait loci (QTLs), Fhb1 and Qfhs.ifa-5A, identifies novel candidate genes. *Mol. Plant Pathol.* **14**, 772–785 (2013).
44. Ge, X., Chen, J., Sun, C. & Cao, K. Preliminary study on the structural basis of the antifungal activity of a rice lipid transfer protein. *Protein Eng.* **16**, 387–390 (2003).
45. Longin, C. F. H. *et al.* Hybrid breeding in autogamous cereals. *Theor. Appl. Genet.* **125**, 1087–1096 (2012).
46. Wu, Y. *et al.* Development of a novel recessive genetic male sterility system for hybrid seed production in maize and other cross-pollinating crops. *Plant Biotechnol. J.* n/a–n/a, <https://doi.org/10.1111/pbi.12477> (2015).
47. Ni, F. *et al.* Wheat *Ms2* encodes for an orphan protein that confers male sterility in grass species. *Nat. Commun.* **8**, 15121 (2017).
48. Xia, C. *et al.* A TRIM insertion in the promoter of *Ms2* causes male sterility in wheat. *Nat. Commun.* **8**, 1–9 (2017).
49. Cigan, A. M. *et al.* Targeted mutagenesis of a conserved anther-expressed P450 gene confers male sterility in monocots. *Plant Biotechnol. J.* **15**, 379–389 (2017).
50. Li, H. & Zhang, D. Biosynthesis of anther cuticle and pollen exine in rice. *Plant Signal. Behav.* **5**, 1121–1123 (2010).
51. Bate, N. & Twell, D. Functional architecture of a late pollen promoter: Pollen-specific transcription is developmentally regulated by multiple stage-specific and co-dependent activator elements. *Plant Mol. Biol.* **37**, 859–869 (1998).
52. Goodstein, D. M. *et al.* Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, 1178–1186 (2012).
53. Leroy, P. *et al.* TriAnnot: A Versatile and High Performance Pipeline for the Automated Annotation of Plant Genomes. *Front. Plant Sci.* **3**, 1–14 (2012).
54. Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **7**(Suppl 1), S10.1–12 (2006).
55. Petersen, T. N., Brunak, S., Von Heijne, G. & Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).

56. Eisenhaber, B. *et al.* Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. *Plant Physiol* **133**, 1691–1701 (2003).
57. Fankhauser, N. & Mäser, P. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics* **21**, 1846–1852 (2005).
58. Kapustin, Y., Souvorov, A., Tatusova, T. & Lipman, D. Splign: Algorithms for computing spliced alignments with identification of paralogs. *Biol. Direct* **3**, 1–13 (2008).
59. Kersey, P. J. *et al.* Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* **46**, 802–808 (2017).
60. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7** (2011).
61. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
62. Letunic, I. & Bork, P. Interactive tree of life (iTOL)v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
63. Hu, B., Jin, J., Guo, A., Zhang, H. & Luo, J. Genome analysis GSDS 2.0: an upgraded gene feature visualization server. **31**, 1296–1297 (2018).
64. Voorrips, R. E. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J. Hered.* **93**, 77–78 (2002).
65. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs\_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics Bioinforma.* **8**, 77–80 (2010).
66. Gaut, B. S., Morton, B. R., McCaig, B. C. & Clegg, M. T. Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcl*. *Proc. Natl. Acad. Sci.* **93**, 10274–10279 (1996).
67. Metsalu, T. & Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**, W566–W570 (2015).
68. Burton, R. A. The Cesa Gene Family of Barley. Quantitative Analysis of Transcripts Reveals Two Groups of Co-Expressed Genes. *Plant Physiol.* **134**, 224–236 (2004).
69. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, 260–266 (2017).
70. Choulet, F. *et al.* Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**, 1249721 (2014).
71. Yang, Z. *et al.* Pistillody mutant reveals key insights into stamen and pistil development in wheat (*Triticum aestivum* L.). *BMC Genomics* **16**, 1–11 (2015).

## Acknowledgements

We are grateful for the support provided by DuPont Pioneer Hi-Bred International Inc. and the University of Adelaide. We thank, Juan Carlos Sanchez and Mathieu Baes for project advice, Dr. Nathan S. Watson-Haigh for his bioinformatics assistance, Yuan Li for technical assistance for the qRT-PCR experiments and, Margaret Pallotta and Dr. Takashi Okada for critical discussions and reading of the manuscript.

## Author Contributions

Conceived and designed the research: A.K., R.W. and U.B. Computational analysis and data collection: A.K., E.K., R.S. and U.B. Contributed to the qRT-PCR experiments: A.K. Analysed the data: A.K., R.W. and U.B. Manuscript preparation and editing: A.K., R.W., R.S., E.K. and U.B.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-35375-7>.

**Competing Interests:** The authors declare 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) 2018